DOCUMENT RESUME

ED 220 279 SE 038 804

AUTHOR	Davis, Robert B.; And Others
TITLE	The Roles of "Understanding" in the Learning of Mathematics. Part II of the Final Report.
INSTITUTION	Illinois Univ., Urbana. Curriculum Lab.
SPONS AGENCY	National Science Foundation, Washington, D.C.
REPORT NO	NSF/SED-82008
PUB DATE	Apr 82
GRANT	NSF-SED-79-12740
NOTE	148p.
EDRS PRICE	MF01/PC06 Plus Postage.
DESCRIPTORS	*Cognitive Processes; Educational Research; Elementary Secondary Education; Higher Education; *Interviews; *Learning Processes; Learning Theories; *Mathematics Education; *Mathematics Instruction; Teaching Methods
IDENTIFIERS	*Mathematics Education Research

ABSTRACT

It is noted that it has long seemed possible to teach mathematics in two distinct ways: (1) learning with understanding, and (2) learning without understanding. The study discussed was designed to ascertain what might be lost if "understanding" was eliminated from instruction, as compared to teaching in which the "understanding" of students was emphasized. An experimental approach was not used in this investigation. Instead, a range of pupils were observed from third grade to calculus students, with episodes identified and categorized that seemed to indicate either a presence or lack of understanding in some particular form. Further, identified behaviors were related to some basic conceptualizations of human information processing that have emerged from recent "cognitive science" studies. (MP)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED220279

NSF/SED-82008

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

● Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

The Roles of "Understanding" in the Learning of Mathematics

Robert B. Davis, Stephen Young, and Patrick McLoughlin

Part II of the Final Report
of National Science Foundation Grant
NSF SED 79-12740

April 1982

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Patricia Babb

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Curriculum Laboratory, University of Illinois, Urbana/Champaign

THE ROLES OF "UNDERSTANDING"
in the
LEARNING OF MATHEMATICS

Robert B. Davis, Stephen Young, and Patrick McLoughlin
Curriculum Laboratory, University of Illinois, Urbana/Champaign

This is Part II of the Final Report of
National Science Foundation Grant
NSF SED 79-12740

It summarizes the substantive results
obtained by this Project; Part I of
the Final Report summarizes administrative
data on the activities of the Project, and
is not of general interest.

April 1982

The Roles of "Understanding"
in the
Learning of Mathematics

Abstract

It has long seemed possible to teach, and to learn, mathematics in either of two distinct ways. One might be called "learning with understanding," and the other might be called "learning without understanding." [In fact, as this study shows, the situation is more complicated than these two simplistic alternatives suggest.] Our society is rapidly changing in a direction that attaches more importance to the various forms of mathematical competence possessed by various groups of people; more effort is being expended to help more people learn more mathematics; new media, such as CAI, micro-computers, calculators, TV, and video discs, are coming into use; and the always-great pressure to operate educational programs as cheaply as possible is becoming even more intense. In such a situation, the prospect looms clear that "understanding" may come to be thought of as an expensive luxury, and be cast aside as non-essential.

Would this entail a substantial loss? The present inquiry, which began on September 1, 1979 and was completed February 28, 1982, sought to examine more closely what might be lost if "understanding" is deemed superfluous -- or, from the opposite perspective, what might be gained if "understanding" of most students could be improved.

What methods should such a study employ? From the traditions of past work in mathematics education, one might expect an experimental approach, with an experimental group taught with emphasis on understanding, their performance then being compared to a group taught without such an emphasis.

In fact, this is not what has been done. Instead, many students -- ranging from third graders working on arithmetical tasks, up to older students who were studying calculus -- have been carefully observed, often

with the aid of tape recordings, and episodes have been identified that seemed to indicate either a particular form of understanding, or to indicate some form of lack of understanding (or the presence of misunderstanding). The goal of these observations was, of course, to obtain a large collection of examples that might serve to delineate the world of "understanding," to help define what it means "to understand" and what it means "not to understand."

As a second activity, these examples were put into various categories to help determine what kinds of understanding seem to be important.

Finally, on a more fundamental level, the behaviors identified in the examples have been related to some of the basic conceptualizations of human information processing that have emerged from various recent "cognitive science" studies.

SECTION ONE

I. The Method of Observation.

Data has been obtained from several sources, including the analysis of student written work (on homework, class "seat work," or tests), the analysis of student performance at computer terminals, the observation of classrooms, observing tutoring sessions, and interviews with teachers and with various categories of "experts," but the main source of data has been the task-based interview. In this procedure, a student (or an "expert") sits at a desk, with paper, pen, and other materials if necessary; an interviewer presents a problem; the student (or expert) attempts to solve the problem, under the agreement that he or she will talk aloud as much as possible, explaining what they are doing, why, what they are thinking about, etc. The interviewer may participate with frequent interjections of questions or other remarks, or may say very little. After the "work" part of this interview is over, the interviewer may ask the subject to explain more fully a few points that seem obscure, or may ask the subject to review the entire episode from memory, adding whatever components seem to deserve mention . The entire episode is usually tape-recorded, and the interviewer usually makes written notes during the session. In some cases there may be an additional observer who also makes notes, as unobtrusively as possible.

The goal is to allow the subject to think about the mathematical task in his or her usual, natural fashion, with as little distortion (from the observation procedure) as possible. Of course distortion does occur, but skillful interview technique seeks to minimize it.

At the end of an interview session, one thus has

i) a taped record of what happened
ii) the subject's written work (which we always have done in ink, to avoid erasures)
iii) the interviewer's notes
iv) the observer's notes, if any
v) the subject's memory of the episode
vi) the interviewer's memory of the episode
vii) the observer's memory of the episode (if an observer was present).

Unavoidably, some interviews take place without tape recording, because mathematical behavior can occur anywhere -- at lunch, at a school athletic event, while walking or driving, at a brief encounter in a school corridor, etc. -- so some of the items on the above list will sometimes not exist.

Where tapes do exist, we do not necessarily transcribe them. The time required to make a good transcription of a 40 minute interview session can be over 50 person-hours of work. Routine transcription of all tapes would not be feasible.

To make matters worse, within our experience only the interviewer or observer can usually make a satisfactory transcription, although typists can often make a first-approximation transcription.

Anyone who studies such tapes carefully will be struck by the fact, hardly mentioned in the literature, that most of the information is NOT coded in the choice of words themselves, and is consequently lost if only words are transcribed. The main message is usually coded in nuance, inflection, timing, pace of remarks, facial expressions, gestures, and posture. Consider, for example, a transcription such as:

(1) Interviewer: ...and what do we have here?
(2) Student: Oh.

The possible meanings of such exchanges are not defined by the mere written transcription of the words.

To be sure, good interview technique can attempt to reduce these uncertainties. [The phrase "...and what do we have here?" was probably a poor choice of words on the part of the interviewer. A better remark might have been:

(1) Interviewer: For the sake of the tape recording, will you please try to say what you just did? [or "...how things now stand?"]

But exchanges of comparable ambiguity continue to occur, in part because the interviewer is constrained to try to distort the subject's normal procedures as little as possible, and must convey the feeling of being interested, non-judgmental, etc.

Perhaps the most trying demands imposed on interviewers are these two:

(i) To use a "poker voice" which does NOT give away clues or cues;

(ii) To avoid "teaching" behavior intended to help the subject out.

Of course, deliberate hints may be given, if they do not interfere with that part of the task which is being studied in that specific interview.

[As the preceding remark suggests, interviews usually have some sort of a priori goal, established in the mind of the interviewer before the interview began. But even in this respect there are trade-offs that must be considered; most of the most important phenomena that we have found were NOT anticipated beforehand -- as, for instance, the surprising inability of students to make drawings or diagrams at the beginning of working on a problem (which we discuss below). Hence, despite the interviewer's initial goals, he must remain open to seeing unexpected things. In particular, he must NOT allow the "efficiency" of his interview procedure to mask important phenomena that he had not anticipated.]

II. The Basic Question

We want to be sure that the reader sees clearly the basic question we are considering. To be sure, as we study the matter we shall see that it is more complex and more subtle than one might at first expect. But the basic question itself -- at least at the outset -- is indeed both simple and important. We look at several instances.

A. One of our studies (Davis and McKnight, 1980) dealt with a third-grade girl, Marcia, who subtracted

$$\begin{array}{r} 7,002 \\ -\ 25 \\ \hline \end{array}$$

by writing

$$\begin{array}{r} \overset{\overset{5}{\cancel{6}}}{\cancel{7}},\ 0^{1}0^{1}2 \\ -\ 2\ 5 \\ \hline 5,\ 0\ 8\ 7 \end{array}$$

Marcia was convinced that she had performed the subtraction correctly. Her teacher's efforts at remediation ultimately failed, and our own efforts at tutoring probably fared no better.

What does this have to do with understanding? Our answer will not be entirely clear until much further along in this paper, because it depends upon a specific conceptualization of human information processing, but we can sketch out the answer now so that the reader will know what to be looking for.

1. First, we would say that Marcia did not understand the size of these numbers. If you or I had, say, a truck or automobile that weighed "about seven thousand pounds," if we removed something that weighed 25 pounds, and if the truck then weighed "about five thousand pounds," we would feel that either a miracle or an error had occurred. Those sizes don't work out correctly. "About seven thousand," minus twenty-five, should still be "about seven thousand." Marcia never responded to arguments of this type; we would say she did not understand the sizes of the numbers. ["Seven thousand" seems to have been merely verbal noise to Marcia, as meaningless as "a trillion" probably is to most American citizens.]

2. Marcia seems not to have understood that "borrowing" and "carrying" operations are her written equivalent of "making change," as in getting ten dimes for one dollar, or ten pennies for one dime. When Marcia wrote

$$\begin{array}{r} \overset{6}{\cancel{7}},\ 0\ 0^{1}2 \\ -\ 2\ 5 \\ \hline \end{array}$$

what she did, in effect, was to trade one one-thousand dollar bill for ten one-dollar bills. She did not see the parallel between such an act and what she wrote on the paper.

3. Interviews with Marcia [all of the details are given in Davis and McKnight, op. cit.] revealed that she was very skillful in representing a numeral (such as 7,002) as an array of Zoltan Dienes' MAB blocks. Use of this representation in this example would have revealed Marcia's error clearly, and also showed her how to correct the error. But Marcia refused to acknowledge that MAB representations had anything to do with this problem!

4. One could summarize much of Marcia's behavior by using the distinction between an algorithm (or specific "recipe"), and an "intuitive idea." [In later sections, we shall define an "intuitive idea" more precisely; thinking of "borrowing" in Marcia's example as "making change" would constitute one relevant "intuitive idea" that Marcia might have used, but did not.]

Roughly speaking, the distinction is as follows: suppose I have some precise instructions written down on a piece of paper, telling me how to find Mr. Wilson's Apple Farm. Suppose that I am attempting to follow these instructions. At some point I encounter trouble. (Perhaps a bridge is out, or I can't find the "red barn" that the instructions tell me to look for.) I ask someone for help. Suppose they tell me my written instructions are wrong, anyhow. What I ought to do, instead, is ...

Do I abandon the written instructions and follow the speaker's advice? Do I try to relate what he is saying to what is written on the paper? Or do I ignore the speaker and keep trying to make the written instructions work? In either of the first two cases I step "outside" of the algorithmic procedure and try to relate the procedure to something else (or even replace it by something else). In the third case I reject the "something else," and insist on following the algorithmic procedure.

Our interviews show that Marcia's behavior was typical. Third and fourth graders are surprisingly devoted to algorithmic behavior, and are reluctant to try to make use of additional "outside" (or non-algorithmic) information. Clearly this is both a strength and a weakness -- if children did not work hard to learn algorithms, present-day school curricula would be even less effective than they presently are. On the other hand, Marcia's commitment to the algorithm she is using stands in the way of her learning to subtract correctly and reliably.

5. There is another sense -- an important one -- in which Marcia doesn't understand. She does not see clearly what she herself is doing. What has made remediation so difficult in Marcia's case is that she believes

i) She has learned the subtraction algorithm carefully and well (and she has, provided there are no zeroes in "inside" columns in the minuend);

ii) She always gets correct answers by using this algorithm (and she does -- again, provided there are no zeroes in "inside" columns in the minuend);

iii) She is using the same algorithm for

$$\begin{array}{r} 7,\ 0\ 0\ 2 \\ -\ 2\ 5 \\ \hline \end{array}$$

that she uses for, say,

1, [illegible] 8 5
- 2 9 6 .

It is, of course, this third belief that causes the trouble. But, unfortunately, one cannot really say whether Marcia is correct, or not, in this belief. There are two possible rules that she might be using:

a) When necessary, "borrow" from the next digit on the left (in the minuend);

or else

b) When necessary, "borrow" from the nearest non-zero digit on the left (in the minuend).

No case had previously arisen that would distinguish between these two rules. (Indeed, the theory of "knowledge" which underlies our work suggests that probably Marcia had not formulated her "rule" so precisely that such a distinction could be described.)

Seymour Papert has emphasized that one reason you need a "teacher" is to *tell you what you are doing*. Imagine a major league home-run hitter who happens to be in a slump. Does he need to have a batting coach tell him how to hit home runs? Of course not; by hypothesis he himself knows that better than anyone else is likely to. Does he need someone to admonish him to stop doing whatever it is that he's doing wrong? Again, of course not. He *wants* to hit home runs again, the way he used to. But -- if he was hitting well then, and poorly now, then *something has changed -- and he himself doesn't know what it is*! He needs someone to help him to analyze his swing, timing, way of looking at pitches, and so on, and try to identify *exactly what has changed*.

[This suggests a very promising line of remediation for Marcia, and for many other students -- *help them to see exactly where their procedure is changing (or where the problem structure is changing), so that a "correct" method suddenly yields some incorrect results*. In all of the literature, which includes abundant documentation of unsuccessful attempts at remediation, we have not found one single instance where this "show them what they are doing" strategy was employed! It would seem well worth trying, especially where other methods are failing.]

B. Erlwanger's "Benny." Erlwanger (1973) used task-based interviews to study the mathematical performance of 5th and 6th graders. One sixth-grader, reported as "Benny," was found to convert $\frac{2}{10}$ to a decimal as 1.2, and to convert other fractions to decimals as follows:

$$\frac{5}{10} \rightarrow 1.5$$

$$\frac{429}{100} \rightarrow 5.29$$

$$\frac{4}{11} \rightarrow 1.5$$

$$\frac{11}{4} \rightarrow 1.5$$

and so on. Erlwanger also found a fifth grade girl who wrote

$$\begin{aligned} .3 + .4 &= .7 \\ 3. + 4. &= 7. \\ .3 + 4. &= .7. \end{aligned} \quad ,$$

putting two decimal points in the same numeral! She had no idea how large .7. actually is -- whether, for instance, it is "more than 6" or "less than 1" -- but this did not strike her as peculiar, inasmuch as she rarely understood the size of the numbers she dealt with on arithmetic papers. (Erlwanger, 1974.)

C. Donald Alderman, Spencer Swinton, and James Braswell (1979) have reported the use of task-based interviews, and written tests, to determine whether fifth-graders in certain (rather typical) schools understood the arithmetic they were learning, in the quite specific sense of being able to make up a meaningful problem to match a mathematical statement such as

$$4 \times 5 = 20.$$

The students were provided with graph paper, with a large collection of cube-shaped wooden blocks, etc. Shading in a 4-by-5 rectangle on the graph paper, or constructing a 4-by-5 rectangular array of blocks, or writing either

$$4 + 4 + 4 + 4 + 4 = 5 \times 4$$

or

$$5 + 5 + 5 + 5 = 5 \times 4$$

-- any one of these -- would have constituted an acceptable answer. (The students might, of course, have gone further, and said "You want to buy four candy bars at five cents each," or "there are 5 school days in a typical week, so in 4 typical weeks there will be twenty days of school" -- but none of them did.) In one class of 24 fifth graders, only 3 could give correct answers. Roughly similar results were obtained when this same task was used in other classes.

D. Dividing Fractions. The sceptical reader can confirm these results for herself or himself. Use this task: "I will show you a mathematical statement, and I want you to tell me some reasonable-sounding story that would correspond to this statement." Now show the subject

$$8 \div 2$$

and explain that a "reasonable-sounding" story might be something like "I have eight dollars in the bank. Every week I withdraw two dollars. How long can this continue?" Or, alternatively, "I have 8 cookies, and you and I will share them equally. How many will you get?"

Do a few practice problems, using only positive integers (including the requirement that the answer must also be a positive integer), to make sure that your subject understands the task itself.

Now, show him

$$\frac{1}{3} \div \frac{1}{2}\ .$$

Unless your subject is professionally-connected to mathematics in some way, it is very unlikely that you will get a correct story. [Note that the story "We have one-third of a pie left. You and I share it equally. How much do you get?" is NOT correct. It corresponds either to

$$\frac{1}{3} \div 2$$

or else to

$$\frac{1}{3} \times \frac{1}{2}\ .]$$

Yet, according to most curriculum schemes, we all learned this topic in the fifth grade, or thereabouts. Of course, most of us did not understand it then, and -- what is far worse -- we did not realize that we *ought* to have understood.

III. What We Are Saying

The background which underlies and motivates this study needs to be distinguished from the study itself. For our specific study of "behaviors that indicate understanding" we have maintained the usual kinds of scientific carefulness. For our discussion of the *background*, by contrast, we make no "scientific" claims. Yet, while maintaining the distinction between the study and the background, it is important to state clearly our view of what that background is. We believe it deals with crucial choices that now face American education.

The background of the study could be described as follows.

A. Arithmetic is typically taught in the United States in what might be described as a stimulus-response (SR) mode. A student sees, say, a flash card displaying

$$4 + 3 = \quad ,$$

and is expected to respond as quickly as possible by saying "seven."

What most of us probably learned in the fifth grade, as far as dividing fractions is concerned, was the admonition to "invert and multiply", i.e.,

$$\frac{2}{7} \div \frac{3}{5} = \frac{2}{7} \times \frac{5}{3} = \frac{10}{21} .$$

But "$\frac{10}{21}$" hardly deserves to be called an "answer," considering that we were probably not at all sure what the question was. It is, of course, a "response."

This form of presentation is usually called "rote" teaching or "rote" learning.

B. If students learn what might be called the "facts" of arithmetic, by meaningless rote, there is very little likelihood that they will use their mathematics -- or be able to use it -- in real situations, which are not typically initiated by a flash-card presentation of symbols like

$$\frac{1}{2} \times \frac{3}{4} .$$

Indeed, real problems are often initiated by something which may not, at first glance, seem at all mathematical.

C. Rote presentations are offensive to many mathematicians (or other heavy users of mathematics), because they know mathematics, and its applications, as a subject where interesting and subtle questions are explored in searching -- and often daring -- ways. What should

$$\frac{1}{2} \times \frac{3}{4}$$

be? What are the logical constraints on the possible ways we might define the product of two positive rational numbers? Which attributes of positive integer multiplication are we most eager to preserve, when we extend the system to the positive rationals? Which meanings do we want to preserve?

D. It could be argued that we are introducing complexities that are inappropriate to the cognitive level of students who are 8, 10, 12, 14, or so years old. Such a claim would be false. There are great differences among students, and some students seem never to reach a condition of serious intellectual curiosity about subtle matters. But many students do. The fact is well established, for example by the collection of filmed and video-taped lessons accumulated by David Page's Arithmetic Project, and by our own Madison Project. These films show actual unrehearsed typical "exploration" lessons, wherein a teacher works with a group of students -- perhaps as young as 8-year-olds -- in the exploration of some mathematical topic, such as the concept of function, or the "square-brackets" function [x], or the concept of isomorphism, or the system of 2-by-2 matrices, or area for shapes on a geoboard, etc. The students are themselves the leaders in these explorations, and the films document clearly how successful the students are, and the quite evident gratification that many students derive from these explorations.

Two decades of "new mathematics" experimentation have left many questions unanswered -- including the basic questions of "What mathematics do we want children to learn?" and "In what sense do we want them to 'know' it?" -- but it can no longer be argued that the process of mathematical inquiry is beyond the cognitive capability, or interest, of 8, 10, or 12-year-old children. Too many films exist that refute any such assertion.

E. It could also be argued that the inappropriateness of rote teaching of mathematics has been widely accepted within the professional literature, at least for several decades. In support of such a claim one could cite Brownell (1945), or Brownell and Sims (1946), or Byers and Herscovics (1977), or a 1940 experiment carried out by Katonah and analyzed by Osgood (1953), or a famous experiment by Bransford and Johnson (1972), among others. Unfortunately, this argument lacks force, for at least two reasons: first, one can also find abundant evidence in the literature that the rote teaching of mathematics has often been accepted as the standard way of doing business; second, no matter what "the literature" shows, direct observation of classrooms indicates that rote teaching is very commonplace in the real world of schools.

F. It could also be argued that many teachers, and many textbooks, do attempt to present mathematics in one or another way that might be called "meaningful." This is clearly true. It is also true that only a small percent of all teachers are involved in such activity -- random selection of classrooms for observation usually fails to turn up any of them, though by good detective work one can find a few such classrooms.

G. Rote teaching of mathematics has been defended on the grounds that it fills an important economic and sociological need -- or did so during the recent past. A few decades ago, this argument runs, industrial societies needed schools primarily as institutions to provide safe, economical custodial care for children. Secondarily, for a sizeable number of students the future held routine work on an assembly line or in an office, where enduring boredom would be important, but originality and creativity would not be. Finally, there was also a need for a quite small percentage of students to master mathematics and physical science profoundly, so as to be able to make important original contributions.

The rote school met these needs ideally. Routine drill and practice provides the cheapest kind of custodial care available, with minimum demands on specialized teacher's expertise, so that substitute teachers can fill in at any time, on a moment's notice. For 90% (or so) of the students, this was exactly what was desired.

For routine assembly line work, or routine office work, a little rote knowledge of arithmetic filled the bill very neatly.

Finally, despite the appalling dullness of the rote curriculum, a few students would nonetheless see through what was taught, glimpse the possibilities of what can really be done with mathematical models of this universe (or of other universes), and become mathematicians, scientists, or engineers. Again, the program would produce just the desired mix of levels of adult performance. Only a few good scientists or engineers were needed.

When the rote curriculum is presented in such a light, it often seems that the writer cannot possibly be serious. There should be no mistake about this. Some writers who see the curriculum this way are entirely serious about their analysis.

The implications should be clear. The meaningless rote curriculum does not come close to matching the socio-economic mix that is needed in the United States today. At this moment, air traffic has been reduced to 80% of earlier levels because of an inability to train enough air traffic controllers quickly enough. Naval vessels have navigational problems (and collisions) that are actually educational problems in the ineffective teaching of vectors to navigational officers. The growth of high-technology industry is being impeded by a shortage of engineers and computer scientists. Japan has become the world's largest producer of automobiles (and may take over computers), thanks to a combination of non-routine work settings, superior education, and a suitable base of personal cultural values. In the U. S., unemployment among some sectors has reached the highest levels since the Great Depression. If the rote-curriculum school matched the socio-economic needs of an earlier United States, it clearly fails -- catastrophically -- to meet the needs of the nation today.

H. One last remark on background: This study had a strikingly "practical" origin. The University of Illinois group were involved in the creation of computer-assisted instruction ("CAI") course-ware, designed to help students learn mathematics in grades 4, 5, and 6. They used task-based interviews, and a study of the literature, to identify weak spots in typical student knowledge or performance, and then lessons were created to meet these very specific needs. For example, in response to Erlwanger's evidence that many students had no adequate idea of the size of the numbers they were working with, Sharon Dugdale and David Kibbey created several CAI lessons dealing specifically with size. One of these lessons was Darts; the terminal's panel displays a picture similar to Figure 1. (The location of the balloons is determined by random numbers, and hence cannot be predicted, nor memorized.) By typing in numbers, the student causes darts to appear at the left of the screen (at the height named by the student's input), to move across the screen from left to right, to thud into the "wall" at the right, and (possibly) to burst a balloon (if the input number matched a balloon's location).

Darts has proved extremely popular with students. The student is left with a number of significant choices in how to burst the balloons. One fifth grade girl, A.C., was observed to apparently waste her first dart, after which every dart burst a balloon. More careful observation revealed that A. C. did not "waste" her first shot -- she used it to get a unit for measuring length! She would type in a number small enough to give her a unit for use in measuring the locations of all of the balloons; thus she might type in (say) $\frac{1}{10}$, then use her fingers to "measure" with the interval $(0, \frac{1}{10})$ so as to locate, precisely, all of the balloons.

FIGURE 1

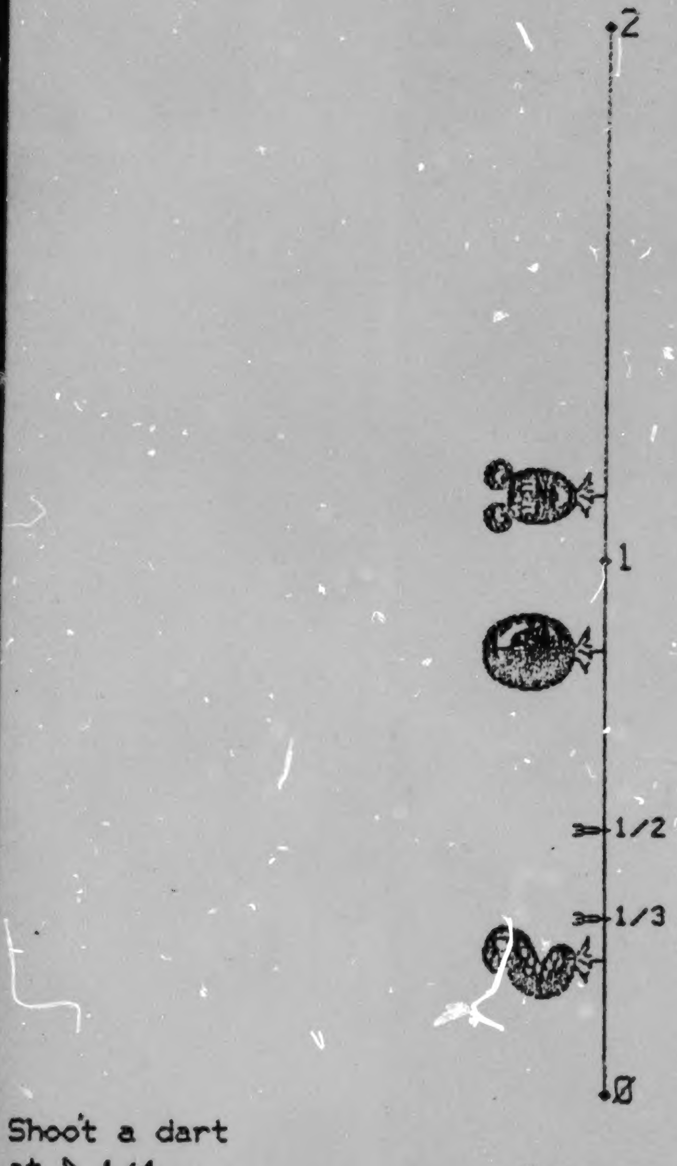


Figure 1 shows that 2 darts have been thrown across the screen: one (following the student's directions) has thudded into the "wall" at 1/2, thereby missing; a second dart, thrown at 1/3, has also missed.

By typing 1/4 at the "arrow" ≫ (Lower Left Corner), the student is telling the computer to throw the next dart at 1/4. The computer will carry out this action as soon as the student presses the "NEXT" key.

OCT 21

Another strategy is to come as close to a balloon as possible, by a "good guess," then correct the height by a systematic correction strategy -- the lesson even allows a student to type in, say,

$$1\frac{5}{7},$$

then $$1\frac{5}{7}-\frac{1}{8},$$

then $$1\frac{5}{7}-\frac{1}{8}-\frac{1}{10}.$$

The lessons produced in this way are known to be successful (Swinton et al., 1978; Davis, Jockusch, and McKnight, 1978), capable of producing important learning gains in students, both in algorithmic performance and in conceptual understanding.

This is the kind of curriculum which we believe is needed. But this belief is not widespread. It is clear that the rote-arithmetic curriculum can be presented, in routine format, by computer CAI lessons, at very little cost in money and effort. Given that possibility, it seems next to certain that the rote curriculum will soon be abundantly available in the form of CAI courseware. Will this amount to an educational program, or to a "classroom pacification program"? We have come to see this as a genuinely urgent question.

IV. Urgency!

One last remark, while standing "outside" of our scientific study itself. The word "urgency" in the previous paragraph does not strike us as excessive. We see the U. S. confronted with an educational system designed and able to teach sterile "facts," but NOT able to teach meanings, nor analytic processes. Given drill-oriented workbooks and drill-oriented computer CAI lessons (or games), this limitation may become more securely built-in in the years ahead.

We do NOT see this problem receiving the attention it urgently deserves.

The resulting failures and dislocations can be at least as bad as our national failure to build automobiles for the years ahead, that devastated the U. S. auto industry in the late 70's and early 80's.

There may well be many reasons still operating to force us, as a nation, to rely so heavily on rote teaching -- excessive class size may be one --but the probable ultimate cost can be horrendous. We must try to analyze this problem from many points of view -- and we must be determined to do something about it!

V. Different *Ways* of "Understanding"

Readers have probably already seen, in our earlier examples, that speaking of "understanding" vs. "not understanding" is an unacceptable over-simplification. Instead, one clearly needs to distinguish alternative ways of understanding. We illustrate this with some observations of eleventh grade high school students who were studying calculus, dealing with applications of the definite integral (Chapter 6 in Anton, 1980). Two classes (16 students in one, 11 in the other), and two teachers, were involved. (Since both teachers were in substantial agreement concerning the forms of understanding that needed to be developed, we report on only one of them.)

A. Calculating the Work Done in Compressing a Spring

The calculation of the work done in compressing a spring is one of the most standard of calculus problems. Three quite different ways to understand this process were revealed from observing these classes:

1. The "Pretend the Force is Constant" Method.

The basic difficulty is this: if the force is constant, the work is computed merely by multiplying force times distance:

$$W = F \times D.$$

Calculus is needed, in the spring problem (and elsewhere) because the force is not constant. Instead, the force depends upon the amount of compression, according to Hooke's Law.

The teacher presented a method for dealing with the non-constant force, as follows:

i) If the force were constant, we'd have no difficulty;

ii) ...but the force depends upon the distance x we have compressed the spring;

iii) ...well, if the distance x didn't change much, the force wouldn't change much...

iv) ...so, arranging things so that x does NOT change much -- just compress the spring a small amount, dx ...

v) and, since the force F will not change much during this small compression (through a distance dx), pretend that the force doesn't change at all. Then one can write

$$dW = F \cdot dx \quad .$$

vi) Of course, in doing this, we have made an error. But we can get estimates on these errors, and we can show that the total error E_n will go to zero when the size of dx goes to zero. (If the total distance the spring is compressed is R, then divide R up into n equal intervals, so that

$$dx = \frac{R}{n} .$$

One then finds that

$$E_n = 0(\frac{1}{n}) .$$)

vii) By taking the limit when $n \rightarrow \infty$, our total error goes to zero, and the summation becomes an integral

$$W = \int_0^R F(x)dx. \qquad (1)$$

2. Define It This Way.

The textbook presents a different analysis of the situation (on page 384), giving equation (1) as the definition of work.

These two approaches represent quite different ways of understanding how work is to be computed. In particular, the "pretend-it's-constant-and-keep-track-of-the-resulting-errors" method has the disadvantage that the language is often confusing, at first, to beginning students, but the advantage that it gives the student a principle that is applicable to many other kinds of problems -- for example, one can compute the force on a non-horizontal bottom of a swimming pool by arguing as follows:

(i) If pressure were constant, one would find the force by merely multiplying pressure times area,

$$F = p \cdot a \quad .$$

(ii) But the pressure is NOT constant; in fact, the pressure depends upon the depth.

(iii) Therefore, we want to break our one "big" problem into n "little" problems, arranged so that the pressure (or depth) is constant (or nearly so) over any single one of the "little" problems;

(iv) ...but this tells us how to arrange the n "little" problems: take narrow strips of the bottom that are all at (nearly) the same depth.

(v) For the rest, proceed as with the spring compression problem.

3. Use the Intermediate Value Theorem.

A third way to understand the work-done-when-a-spring-is-compressed problem is to observe that the force, F, is a monotonically (indeed, linearly) increasing function of the compression distance, x. Using, for each sub-interval, the smallest value of x will thus produce an approximation for F which is too small; using the opposite end-point of each x sub-interval would produce an approximation which is too large. Thus, by the Intermediate Value Theorem (Anton, p. 185), there is a value x_k^* in each interval (x_k, x_{k+1}) such that $F(x_k^*)(x_{k+1} - x_k)$ gives a correct value of the work ΔW_k, with the error equal to zero. But, by the definition of the Riemann integral,

$$\lim_{n \to \infty} \sum_{k=0}^{n} F(x_k^*)(x_{k+1} - x_k)$$

is precisely

$$\int_0^R F(x)\,dx.$$

These constitute three quite different ways to "understand" the computation of the work done when a spring is compressed.

4. Multiple Understandings -- "All of the Above."

The existence of three different ways to understand work automatically implies at least a fourth: one can "understand" all of the others, and see them as legitimate alternatives -- indeed, see how they relate to one another, see their various advantages and disadvantages.

Seeing several possible understandings is significantly different from seeing only one, no matter which that one may be.

B. Using Formulas vs. "Cans of Tomato Soup".

One can argue that the matters referred to in A, above, are somewhat exotic, and lie at the outer limits of most student's cognitive awareness. The distinction we consider here, however, was well within the thinking of these eleventh-graders.

The textbook deals with volumes of solids of revolution by presenting a few basic formulas, such as

$$V = \int_a^b \pi\,[f(x)]^2\,dx$$

[Anton, p. 362]

and

$$V = \int_c^d \pi\,[u(y)]^2\,dy,$$

Initially, nearly all students in both classes dealt with such problems by merely substituting into these formulas. Both teachers, independently, decided that this was a poor way to deal with such problems, and sought to get students to realize that one did NOT need to memorize (or use) these formulas. Instead, one could think of a small volume, dV, shaped like a cylinder with a very small height. Both teachers spent considerable time trying to ensure that students overcame the probable limitations on what they think of as "cylinders." A can of tomato soup, a Necco wafer, and a dime are (neglecting minor variations such as flanges, milling , writing, etc.) all cylinders. In a mathematical sense, a dime is the same shape as a can of tomato soup. To be sure, for the soup the radius may be 1.5 inches, or so, and the height may be 4 or 5 inches, or thereabouts, whereas for the dime the radius is perhaps $\frac{1}{2}$ cm., and the height (or thickness) is perhaps 1.5 mm. -- but, ornamentation and details aside, both cans and dimes are cylinders. For any cylinder -- even one with very small "height" (i.e., thickness) -- the volume is

$$V = \pi r^2 h \quad ,$$

the area of a face, times the height.

Some students had trouble at first seeing a circular disc, cut from a sheet of paper, as a "cylinder" -- the "height" (thickness) seemed too small to qualify -- but once this was accepted, it became possible to write

$$dV = \pi r^2 \, dx,$$

to identify "r" and "x" (or "y") in any particular problem, and thence to find the volume by integration.

The contrast, then, was between a first method:

(i) substituting into formulas such as

$$V = \int_a^b \pi \, [f(x)]^2 \, dx$$

vs. an alternative method:

(ii) computing dV by recognizing the small "slice" in question as a cylinder (or as some other well-known simple shape).

Both teachers advised students to use the second method, tried to make this easier by emphasizing the true shapes (as "cylinders", etc.) of the various tiny slices, and enforced this choice by putting on tests problems for which the textbook did NOT give formulas.

Summary: one thus sees two quite different ways of "understanding" the methods of finding certain volumes by integration. (The teachers were consistent, both here and in the "work" problems discussed earlier, in trying to make intuitive sense out of separate "little pieces," such as

$$\Delta W_k = F_k \Delta x_k$$

and

$$dV = \pi r^2 \, dx.$$

[One could quarrel over whether this is better written as "ΔV" and "Δx."])

C. Arc-length vs. area.

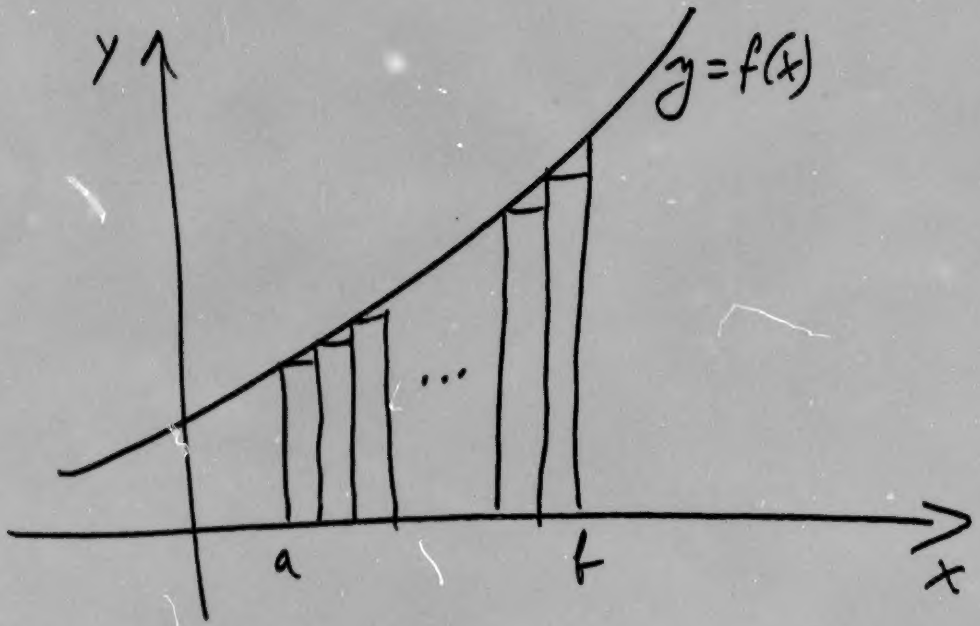
The multiplicity of different ways of "understanding" becomes considerably greater when one considers arc-length [Anton, p.374ff]. Again, one way to deal with such problems was to use the formula

$$L = \int_a^b \sqrt{1 + [f'(x)]^2} \, dx$$

(Anton, p.375).

Both teachers were again consistent, and discouraged this. The "understanding" which one teacher sought to build in students' minds was roughly as follows:

(i) It is important to notice how arc-length differs from area. In dealing with area, we used inscribed rectangles,



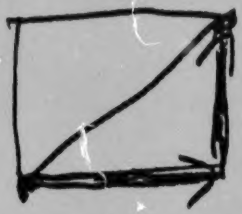
and thus were faced with a "stair-shaped" approximation to the smooth curve y = f(x);

(ii) For each rectangle, the error is a small somewhat "triangular" error that is easily shown to be $0\left(\frac{1}{n^2}\right)$. The sum of such errors is thus $0\left(\frac{1}{n}\right)$, and becomes small when n becomes large.

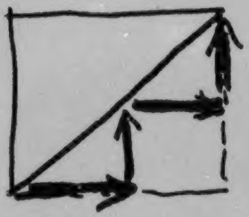
(iii) Notice that it would NOT be enough to show that the errors for individual "stair-steps" goes to zero, because so does the area of the rectangle itself. On every hand, we are dealing with a very large number of very tiny numbers. But the total error goes to zero, whereas the sum of the areas of the rectangles approaches the correct value for the area under the curve.

(iv) Now, arc-length behaves very differently. The teacher demonstrated this by using a "discovery" of a twelve-year-old student at the school. Consider the distance from (0,0) to (1,1). Clearly, this is $\sqrt{2} \doteq 1.414$. But suppose you go from (0,0) to (1,0), and thence to (1,1). The distance is now 2. Suppose you modify the path like this:

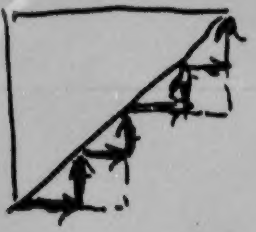
change



to be ,

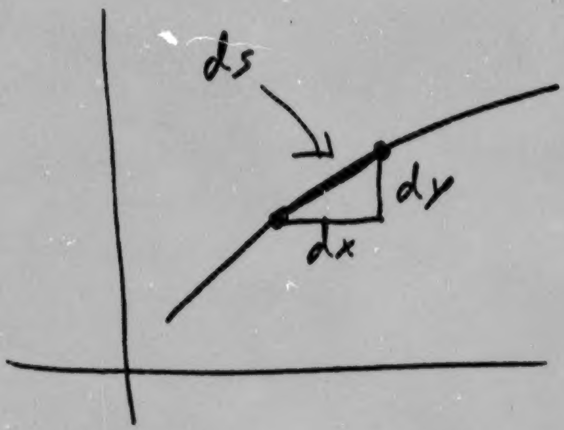


then change it again, to get



and so on. You can "get very close to" the straight-line path from (0,0) to (1,1); but at every step, the "stair-way" path has length exactly two. "Taking smaller sub-problems" isn't decreasing the error at all!

(v) Something quite different is required. Indeed, one needs to use the hypotenuse of a tiny right triangle



and to use the familiar result that

$$ds^2 = dx^2 + dy^2.$$

(vi) The teachers wanted students to see how to use this idea to find arc-length if one is given

$$y = f(x),$$

or if one is given

$$x = g(y),$$

or if one is given both x and y as functions of a parameter (say, t):

$$x = f_1(t)$$

$$y = f_2(t).$$

(vii) The teacher also wanted students to see that this kind of understanding led to greater flexibility. For example, given this, with some careful thought the students could find the arc-length of a helical spring:

$$x = \sin \theta$$
$$y = \cos \theta$$
$$z = \theta \ ,$$

since they could work out for themselves the relationship

$$ds^2 = dx^2 + dy^2 + dz^2 \quad ,$$

by repeated applications of the Theorem of Pythagoras.

(viii) From earlier work on limits, the teacher wanted the students to master at least three views of the limit process:

(a) An "algebraic" or "algorithmic" view, obtaining formulas such as the algebraic estimate on the sum of the errors;

(b) Some notion -- not yet formalized -- about "taking the limit" when "n goes to infinity." [This, of course, became quite a large topic in its own right.]

(c) A realization that these various symbols really refer to numerical quantities -- to numbers! Students considered the effects of sums and products of "ordinary-sized numbers" [like 1 or 7 or 1/2], "very small numbers" like

.00032] and "infinitessimals of higher order" [like $(.00032)^2 = .000000102$], with emphasis on sums such as

$$.00032 + (.00032)^2 = .000320102,$$

$$\begin{aligned} &.00000032 + (.00000032)^2 \\ &= .00000032 + .0000000000001024 \\ &= .0000003\underbrace{20000}1024 \end{aligned}$$

Six correct significant figures!

The similarity between this and Marcia's subtraction difficulties should be obvious. Marcia possessed the ability to represent decimal numerals as arrays of MAB blocks, and could "trade" MAB blocks correctly. If she had made use of this knowledge, it should have caused her to recognize the error in

$$\begin{array}{r} \overset{\overset{5}{\cancel{6}}}{\cancel{7}},\ 0\,{}^{1}0\,{}^{1}2 \\ -\quad 2\ \ 5 \\ \hline 5,\ 0\ \ 8\ \ 7 \end{array} \quad ,$$

and should even show her how to correct the error. But this did not occur. Even after prompting from the interviewer, Marcia claimed not to see the relevance of MAB blocks to the subtraction problem she was working on.

Similarly, the calculus students frequently did not see that

$$\text{Total error} = 0\left(\frac{1}{n}\right)$$

had anything to do with numbers like

$$.00000032 + (.00000032)^2 = .0000003200001024,$$

that finding volumes from slices had anything to do with the formula

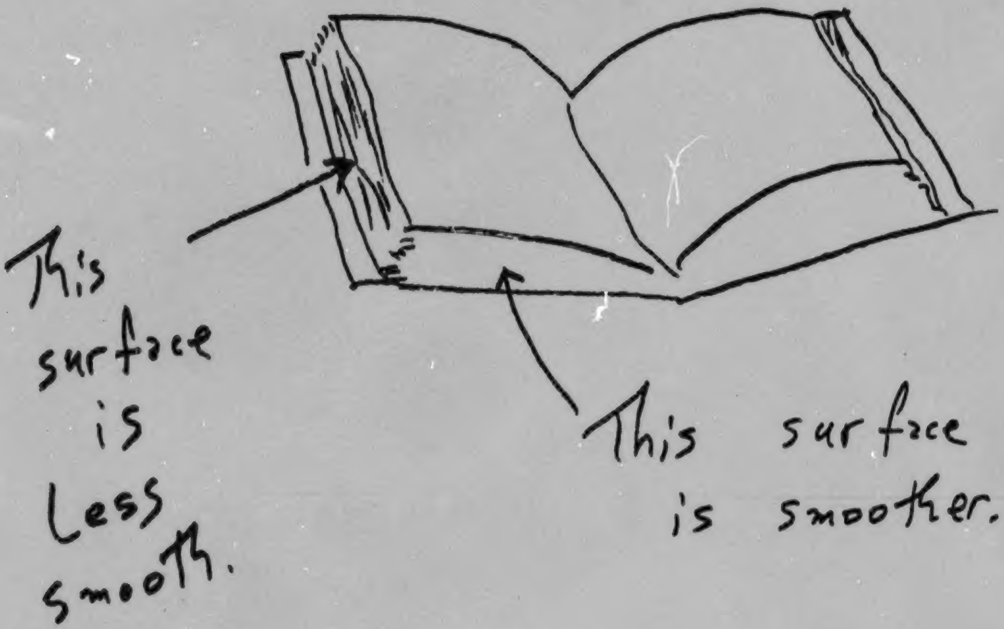
$$V = \pi r^2 h$$

for the volume of a cylinder, or that finding arc length had anything to do with the Theorem of Pythagoras. Both teachers worked continually to establish these connections in the students' minds.

D. Surface Area vs. Volume.

The three views of limits listed above -- "algebra," "limits," and "numbers" -- were extended by the teacher to include a fourth, when the discussion moved on to three-dimensions. To appreciate what was done, it is important to notice that what, in two-dimensions, is arc-length will appear, in 3-D, as surface area; and what in 2-D is area will become, in 3-D, volume. Thus, in 3-D, the presence of "stair-step" indentations (rather like the sides of Egyptian pyramids) will not destroy the correctness of the volume calculation, because the total error caused by the indentations will go to zero as the step-size goes to zero. For the surface area this will not happen. (The "diagonal of the square" counterexample still applies, if a third "z" dimension is added.)

As a fourth way of thinking about these phenomena, the teacher used physical objects. The calculus book itself provides a typical example: if the book is lying open on a table, one can feel a difference between the smoothness of the bottom-of-the-pages side of the book, vs. the perceptibly-greater roughness of the side-of-the-pages side.


This surface is Less smooth.
This surface is smoother.

This is, again, the "steps-on-the-side-of-an-Egyptian-pyramid" phenomenon, where the step size is just the thickness of a single page of paper.

The teacher was concerned that the students be able to see how the

$$V = \int dV$$

calculation relates to physical objects and physical situations; hence he developed four approaches to the definite integral:

(a) an "algebraic" calculational approach;
(b) use of an intuitive theory of limits;
(c) an "arithmetic" view, looking at the comparative size of some of the numbers;
(d) meanings of dV, dS, etc., in actual physical situations with actual physical objects.

These represent four different ways of "understanding" the uses of the definite integral; they thus imply a fifth way: understanding all four, and the relationships between them.

E. "Using Formulas" vs. "Cutting Paper Bands"

The earlier distinction of "substituting into formulas" vs. "recognizing familiar objects" (such as wafer-thin "cylinders") -- cf. Section B, above -- appears throughout calculus. In the case of finding the area of a surface of revolution, one can use the formula

$$S = \int_a^b 2\pi f(x)\sqrt{1 + [f'(x)]^2}\, dx ,$$

or one can recognize a thin circular band. If this band is cut across its narrowest dimension, and flattened out, it becomes (except for infinitessimals of a higher order) a rectangular parallelepiped -- intuitively, "brick-shaped." The dimensions of this brick -- "height," "length," and "width" -- are now readily recognizable, so that

$$dA = 2\pi r\, ds,$$

with r and ds easily determined. [The teacher took pains to make sure that the students realized that a piece of paper, a pad of paper, and a brick are all the same shape.]

Again, there are at least two quite different ways to "understand" this topic.

In fact, all of the calculus examples discussed briefly thusfar actually involve intricate arrays of details. A more complete discussion might profitably employ tree diagrams to chart the paths through alternative forms of "understanding." But our main point should be clear: with so many, quite different ways of thinking about mathematical problems, it is clearly NOT adequate to contrast "understanding" with "not understanding." One must go further and ask: 'In what way, precisely, does this student "understand" this topic?'

VI. Arithmetic vs. Calculus

The range from arithmetic to calculus may seem extreme, and therefore perhaps inappropriate for a single study. We realize that some readers will find it uncongenial, being interested in one end of the range but not the other. Yet the overall patterns of what it means "to understand" are strikingly similar at both ends, and everywhere in between.

We acknowledge that "understanding" in calculus reaches into vast areas of relevant -- even *essential* -- knowledge. One cannot understand calculus without understanding *limits*. One probably should not be said to "understand" limits unless one has both a formal understanding and also an intuitive understanding. Neither of these is possible unless one knows quite a few examples and counter-examples. One must know the topology of the real line, both formally and intuitively. One must understand mathematical induction and proof by contradiction. And throughout all of this, algebraic calculational skill is required in order to fit the pieces together. One can look more deeply into heuristic problem-solving skills in calculus.

This is so large an area that we can report here only on pieces of this vast territory. But even the pieces are interesting.

SECTION TWO

VII. What Does It Mean to "Understand"?

Our serious answer to this question is presented in Section Four, because it depends upon a conceptualization of thought processes which we develop in Section Three. In this section we give what might be called some "naive" answers to this question. In some cases we will build on positive examples, and in others on negative instances, since either can sometimes clarify the meaning of "understanding."

A. Sometimes we "understand" because we are able to match a specific input to something that we can retrieve from memory, and find

(i) the match is perfect

(ii) the "something" that we retrieved from memory leads to associations that answer all of our present needs.

Example: The equation

$$e^{2t} - 5e^{t} + 6 = 0$$

is easily dealt with if, first, we retrieve the general quadratic equation

$$ax^2 + b\,x + c = 0$$

and its associated "quadratic formula"

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

and make use of the perfect match

$$\begin{aligned} e^{t} &\longleftrightarrow x \\ e^{2t} &\longleftrightarrow x^2 \\ 1 &\longleftrightarrow a \\ -5 &\longleftrightarrow b \\ 6 &\longleftrightarrow c \quad , \end{aligned}$$

whence

$$x = 2$$

or

$$x = 3 \ ,$$

so that

$$e^t = 2$$

or else

$$e^t = 3.$$

Now this process is used again: we retrieve from memory some knowledge about logs and exponentials being inverse functions.

$$A^r = s \quad \Longleftrightarrow \quad \text{Log}_A \ s = r.$$

We again find a perfect match between our input data and our retrieval from memory

$$\begin{aligned} e &\longleftrightarrow A \\ t &\longleftrightarrow r \\ 2 &\longleftrightarrow s \quad , \end{aligned}$$

so that

$$t = \ln 2$$

or (similarly)

$$5 = \ln 3.$$

The kind of "perfect match" has been elegantly described by Minsky and Papert (1972), who say that when the key parts of a situation on a chess board has been matched perfectly to a retrieved piece of "knowledge" -- say, perhaps, we recognize that a knight is pinned by a potential queen attack on the king -- "it is almost as if the pieces involved suddenly changed color." Suddenly we understand what the knight can, and cannot, do -- and what this implies for the mobility of other pieces (Davis, 1982-A.) What might previously have been many tiny pieces of separate input data has been reorganized into a larger "chunk" (in George Miller's phrase). It is a single thing, and this single thing is connected to many other items stored in our memory. A similar view of "meaning" is presented in Hofstadter (1980).

This is a useful naive category of "understanding," but it leaves much that requires further discussion.

B. Sometimes we fail to retrieve something from memory, even though the item was stored in memory. This is one way in which we can fail to "understand." An example was given earlier, when Marcia failed to retrieve key pieces of knowledge which she did possess.

A failure of understanding, then does not necessarily imply that there was no knowledge in memory which could have been helpful -- it may be only the retrieval process that has failed.

C. In case A we considered a "perfect match" between input data and some knowledge representation retrieved from memory. When successful, this can constitute a powerful kind of "understanding." But it can fail in any of several ways.

1. In the first place, the retrieved knowledge representation needs, like Janus, to look in two directions. It must accept the input data, and in that sense it must look at, or connect with, the present specific problem. But it must relate also to knowledge previously stored in memory which is relevant to the present problem. If this other knowledge is not stored in memory, or if the retrieved representation fails to connect to it, then one will "not really understand."

Example: a class of 16 eleventh-graders studying calculus. In "clock arithmetic" on an American 12-hour clock, "12" is really a name for the additive identity element, and might better be called "0". Then one has "divisors of zero" -- e.g., 2 x 6 = 0 (although 2 ≠ 0 and 6 ≠ 0), or 3 x 4 = 0 (although 3 ≠ 0 and 4 ≠ 0). Asked where in high school mathematics non-existence of divisors of zero (for certain number systems) played an important role, the students could think of no such situation. [A correct answer could be: in solving polynomial equations by factoring.]

We have found repeatedly that the axiom $A \times B = 0 \Rightarrow [A = 0 \vee B = 0]$ does not seem to be associated with the algorithms which depend upon it, despite teacher emphasis to try to achieve such an association (Cf., e.g., Davis, Jockusch, and McKnight, 1978).

In fact, our earlier example with Marcia is probably an instance of this same phenomenon: the representation structure which is "active" at the moment does not connect with other knowledge which should be seen as relevant.

2. Has the correct representation structure been retrieved from memory?

"Retrieval-and-matching" frequently fails because the wrong piece of knowledge has been retrieved.

With the calculus students we observed, this frequently occurred in attempts to solve quadratic equations. The students, assuming that they were dealing with a linear equation, would try to pursue the strategy of "getting all the x-terms on the left of the equals sign" (although they often did this incorrectly, as in

$$x - 20 + \frac{96}{x} = 0$$

$$x = 20 - \frac{96}{x} \qquad).$$

In such cases we see a "failure of understanding" that consists essentially of retrieving the wrong piece of knowledge from memory, not recognizing the error, and attempting to match the input data with this wrongly-chosen knowledge representation structure.[1]

---

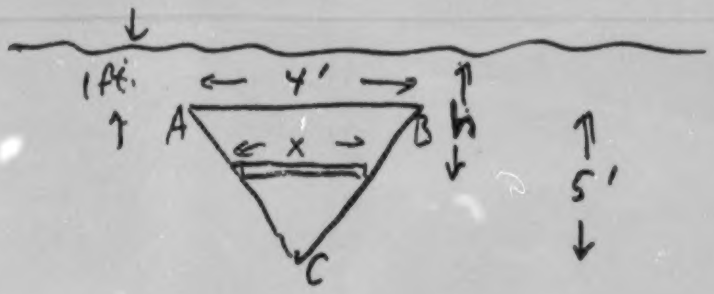
[1] We shall say more about "knowledge representation structures," "frames," "scripts," etc., in Section Three, below. See also Davis, 1982-B.

3. Incorrect mapping into a correctly-chosen representation.

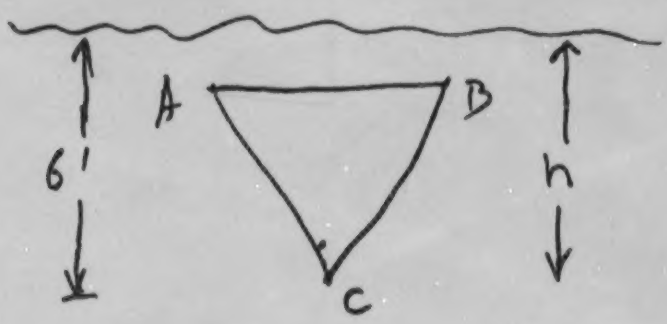
Kirsten, an eleventh-grader, was working on this problem:

> A triangular plate ABC is submerged in water with its plane vertical. The side AB, 4 ft. long, is one foot below the surface, while C is 5 feet below AB. Find the total force on one face of the plate.

Kirsten, not following the teacher's recommendation, used variable names in her diagram to label "end-point" dimensions. The teacher had recommended drawing variables at intermediate values



as here, where h is shown more than 1 foot and less than 6 feet, Kirsten, however, made this diagram



so that h was shown at its maximum value, 6 feet.

A key step in solving the problem is to get x as a function of h. For the correct relation, we have, from similar triangles,

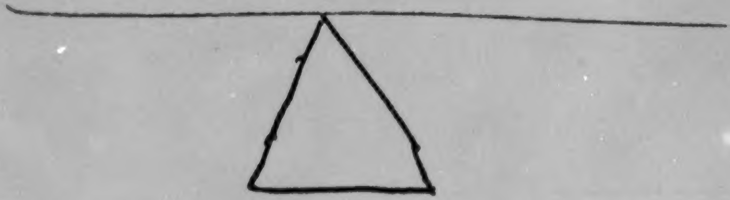
$$\frac{x}{4} = \frac{6 - h}{5} \quad .$$

Kirsten realized that x would be a linear function of h, and wrote

$$x = ah + b,$$

and concluded that $b = 0$, and $a = \frac{4}{5}$, so that $x = \frac{4}{5}h$.

The teacher tried to show Kirsten that this meant that x would increase as h increased, which would imply a geometric configuration generally like



No argument reached Kirsten; she was convinced that everything she had done was correct.

This insistent belief in the correctness of a wrong solution has been reported frequently in the literature in recent years (see, e.g. Rosnick and Clement, 1980 ; Davis, 1980-B). It has to be recognized as one of the most provocative and revealing phenomena reported in decades. [Cf. also Davis and McKnight (1980) and Davis (1982-B)].

Kirsten shows this pattern repeatedly, and is sometimes literally reduced to tears by the intolerable frustration of "correct" work somehow unaccountably going astray.

Where has Kirsten made errors? Her picture is essentially correct, except for showing h at its extreme value, making it hard to distinguish between the constant function $h(x) = 6$ for all x, as against a variable h that covers the interval [1,6],

$$1 \leq h \leq 6.$$

This probably makes it harder for Kirsten to analyze the variation in x, and in h, and the relationship between them.

Kirsten is also correct in thinking that h is a linear function of x, and conversely:

$$ax + bh + c = 0$$

She is correct in concluding that a change of 5 in h corresponds to a change of 4 in x. Putting Kirsten's remarkable persistence into context with other similar studies (including the case of Marcia), it seems likely that the correctness of much of what she has done is standing as an obstacle to her seeing that there is an error in her work. As she checks over her reasoning, it seems to her that she has retrieved the correct tools from memory (she has), and that she has mapped the specific present input data into the appropriate slots in a correct way (she has not!).

Whatever the details of this information processing, this kind of error is typical of Kirsten. On another occasion, Kirsten tried to solve this problem:

Find $\frac{dy}{dx}$, for $x = \frac{\pi}{12}$ and $y = 3$,

if

$$\sqrt{3 + \tan xy} - 2 = 0.$$

She was unable to, and finally decided she wanted to watch while the teacher solved the problem.

The teacher wrote:

$$\sqrt{3 + \tan xy} - 2 = 0 \qquad (1)$$

$$\sqrt{3 + \tan xy} = 2 \qquad (2)$$

$$3 + \tan xy = 4 \qquad (3)$$

$$\tan xy = 1 \qquad (4)$$

$$(\sec^2 xy)\left(x\frac{dy}{dx} + y\right) = 0 \qquad (5)$$

$$x\frac{dy}{dx} + y = 0 \qquad (6)$$

$$\frac{dy}{dx} = -\frac{y}{x} = -\frac{3}{\frac{\pi}{12}} = -\frac{36}{\pi} .$$

Here is an excerpt of the discussion at this point:

Kirsten: What happened to the secant squared?

Teacher: Well, the secant squared is never less than one -- so it's never equal to zero -- so, by the Zero Product Principle [Stein and Crabill, 1972], the other factor must be zero, and that's where I got this equation [pointing to equation 6]...

Kirsten: But where did the secant squared [heavy emphasis] go? You didn't do anything with the secant squared!

Teacher: Kirsten, suppose I have two numbers, A and B, and suppose I know that

$$AB = 0.$$

What do I know about these numbers?

Kirsten: One of them has to be zero. [K. says this very quickly, glibly, her manner dismissing this as irrelevant]. I know all of that! But what happened to the secant squared? [Again, heavy stress on these last two words.] [Note: once again, Kirsten is close to tears.]

Teacher: Kirsten, look -- this [$\sec^2 xy$] is a number, and this [$(x\frac{dy}{dx} + y)$] is a number, and if we multiply these two numbers together, we know we get zero. Now, this number

$[\sec^2 xy]$ *cannot* be zero. Therefore, *this* number

$$x \frac{dy}{dx} + y$$

must be zero, and that's what I've written.

Kirsten: But what happened to the *secant squared*? You've just dropped the *secant squared*!

Something has to be going on here. Two experienced teachers interpreted this episode differently. The first thought Kirsten had retrieved the familiar algorithmic recipe used, say, in

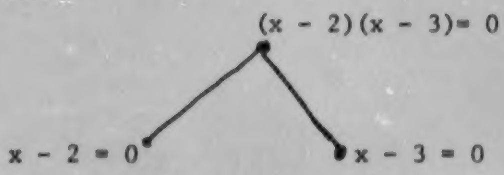
$$(x - 2)\,(x - 3) = 0$$

$\therefore$ Either $x - 2 = 0$, or else $x - 3 = 0$.

If $x - 2 = 0$, then $x = 2$.

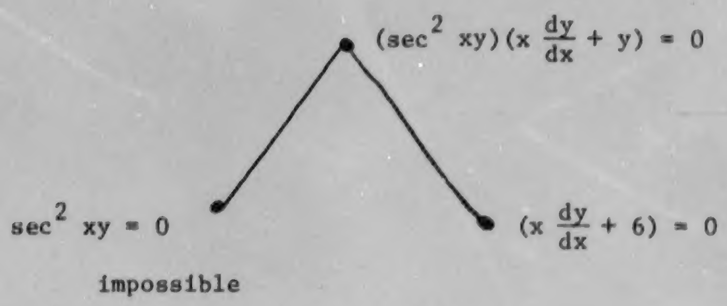
If $x - 3 = 0$, then $x = 3$

which we can represent as a tree



Kirsten followed this algorithmically (much as Marcia did in an earlier example), and would not interfere with the execution of an algorithm by allowing the intrusion of other relevant semantic information.

The other teacher thought that Kirsten had mapped this specific problem correctly into the Zero-Product-Principle decision tree, but had not developed the meta-language necessary to allow her to talk about this algorithm. This teacher thought that the interviewer had made an error in not following out both branches of the decision tree :


$(\sec^2 xy)(x \frac{dy}{dx} + y) = 0$
$\sec^2 xy = 0$
impossible
$(x \frac{dy}{dx} + 6) = 0$

When the interviewer tried to get Kirsten to trim the tree by deleting one branch, he lost her. She could not carry out this kind of meta-analysis. Had the interviewer "gone through all the motions" of following out each branch, perhaps Kirsten would have "understood."

This suggests the following additional kind of understanding (or of misunderstanding):

D. Possession of (and Use of) an Adequate Meta-Language

Kirsten, above, is an example. Although she could easily execute the Zero-Product-Principle algorithm, she was seemingly unable to talk about it, and could not "understand" what the teacher was saying.

E. Clearly, "understanding" cannot be 100% retrieval of relevant ideas learned on previous occasions. Such a system could deal with nothing new -- but people do deal successfully with new inputs, and do so every day! Thus, "understanding" must, at least in part, frequently mean the ability to construct a mental representation for the present input data.

But a given person may not possess the capability of creating a representation for some specific input data. We shall consider this in more detail in later sections. For the present, consider the example of "understanding" the gigantic molecular clouds which astronomers study:

(i) These clouds are the most massive objects in the galaxy.

(ii) One such cloud may have a mass that is 200,000 times the mass of the sun.

(iii) Yet these clouds are almost "perfectly empty" space. Each of these clouds is a more perfect vacuum than any that has ever been produced on earch (Blitz, 1982).

Do you have a clear picture of what one of these clouds is like? Mathematically-sophisticated readers may be beginning to create a reasonable representation in their minds -- if they did not previously have one -- but less sophisticated readers perhaps cannot do so.

But it gets worse:

(iv) The clouds are very "lumpy" -- within a cloud, the gas is organized into "clumps" -- within each clump, the density of the gas is ten times greater than the average density for the entire cloud.

(v) How large is one such cloud? A typical dimension across one cloud would be about 45 times as great as the distance from our sun to the next star nearest to our sun, but a dimension across one cloud might in some cases be more than twice that large.

(vi) These large clouds may exert gravitational and tidal forces on stars, or on star clusters. These forces may be large enough to disrupt less-stable clusters of stars. [Quite a feat for an almost-perfect vacuum!]

(vii) Stars are often created inside a giant cloud, when (in some complex way) some of the matter comes closer and closer together. If a star is born inside a giant cloud, if it moves with a speed of 10 kilometers per second with respect to the cloud, and if the star lives for three million years before it "dies" or "self-destructs", it may still be in that same cloud when it dies! (That gives some idea of how large these clouds really are...)

Can you represent all of that information adequately in your own mind? This example is interesting because the separate ideas -- speed, distance, etc. -- are things we all know. Hence this is NOT a case where we don't know the usual meanings of the words. What is difficult is to fit all of the parts together so as to get an adequate and coherent representation. A major cause of the difficulty lies in the size of some of the quantities that are involved. We are not, for example, accustomed to thinking of a gas so thin that it is a more perfect vacuum(!) than any ever created on earth, yet at the same time more massive than our sun -- indeed, more than 200,000 times more massive than the sun! But then, we do not often think of distances that are 45 times the distance from the sun to the next nearest stars or possibly 100 times that distance. Such a thin cloud can be so massive, clearly, because it is so big!

Far more complicated -- and self-contradictory -- examples exist within high school mathematics, as we shall see below.

"Understanding," then, is sometimes closely related to the ability to construct a suitable mental representation in your own mind.

For some mysterious reason, relatively little attention seems to have been paid to the task of taking specific input data and using it as a basis for the creation of a mental representation of the problem situation -- yet our observations indicate that, among the students we have observed, this may be the point where failures are most likely to occur.

F. Seeing the Story Line.

In fact, the solution of a mathematical problem is very similar to a chess game. Every step that is taken serves some purpose. One sense of "understanding," then, is setting appropriate sub-goals, or seeing the reason why a step is taken.

Within our study, one sub-study consisted of asking 27 eleventh graders, who were studying calculus, to find

$$\lim_{x \to 0} \frac{\sin 7x}{\sin 3x} ,$$

with the requirement that they justify each step of their work by reference to the known limit

$$\lim_{\theta \to 0} \frac{\sin \theta}{\theta} = 1 .$$

What this task amounts to is a shrewd setting of sub-goals, for example:

$$\lim_{x \to 0} \frac{\sin 7x}{\sin 3x} = \lim_{x \to 0} \frac{\sin 7x}{7x} \cdot \frac{7x}{\sin 3x}$$

[which gives us one opportunity to use the known limit]

$$= 7 \lim_{x \to 0} \frac{\sin 7x}{7x} \cdot \frac{x}{\sin 3x}$$

[which reduces clutter]

$$= \frac{7}{3} \lim_{x \to 0} \frac{\sin 7x}{7x} \cdot \frac{3x}{\sin 3x}$$

$$= \frac{7}{3} \cdot 1 \cdot 1 = \frac{7}{3} .$$

Alternatively, one could set a different sequence of sub-goals, as in

$$\lim_{x \to 0} \frac{\sin 7x}{\sin 3x} = \lim_{x \to 0} \frac{\frac{\sin 7x}{x}}{\frac{\sin 3x}{x}}$$

$$= \lim_{x \to 0} \frac{7 \cdot \frac{\sin 7x}{7x}}{3 \cdot \frac{\sin 3x}{3x}}$$

$$= \frac{7}{3} \lim_{x \to 0} \frac{\frac{\sin 7x}{7x}}{\frac{\sin 3x}{3x}}$$

$$= \frac{7}{3} \quad .$$

Many other possible sequences can be created.

Of the 27 students in our study, only one was able to solve this problem correctly -- the youngest student in our study, who was 14 years old at the time, and who has consistently been the best mathematician among the students we have been studying.

Of course, all 27 students could easily deal with such problems AFTER seeing one solved correctly. What we were discussing above was the ability to tackle this as a novel problem, to set up appropriate sub-goals, and then to carry out this plan.

G. There is a special case of sub-goals that deserves an explicit listing. "Understanding" any long proof or calculation seems to be impossible, unless one can break the argument up into sections that are joined together to make the complete proof or calculation. (Note 1)

H. We have seen (in the case of Marcia and her use of subtraction algorithms) that "understanding" can also be a matter of "understanding what you yourself are actually doing."

I. Understanding the Nature of the Task.

Anne, a nineteen-year-old student at a community college, was studying Laplace transforms, and came to us asking for help. Her central difficulty, it became clear, was that she did not understand the kind of thing she was working on.

The idea of Laplace transforms [cf., e.g. Hildebrand (1949), or Finizio and Ladas (1982)] is this: one wishes to study functions f (x) described (say) by an initial value problem for a linear differential equation, such as

$$\begin{cases} y'' + ky = e^{ax} \\ y(0) = {}^{-}1 \end{cases} .$$

Instead of dealing in the space S of functions $y \equiv f(x)$, where differentiation and integration are the key processes, one can map the functions in S into a different collection of functions, L, and carry out the work in L. Why would one do this? Because the key operations in L are very simple algebraic operations. The operation of differentiation in S corresponds to multiplication by s in the space L. Integration in S corresponds to division by s in space L.

Readers familiar with the computational uses of logarithms (a topic now obsoleted by hand-held calculators) will recognize a very close parallel to the use of logarithms to replace multiplication problems by addition problems, using

$$\log AB = \log A + \log B.$$

(Other examples of using mappings or isomorphisms could be cited, as well.)

Anne did not know of this view of Laplace transforms, and was lost in a maze of formulas for which she could see no purpose.

From compiling a "naive" list of kinds of understandings, and looking at examples, it became clear that a more explicit use of a more precise conceptualization of human information processing would be needed if one were seeking to understand "understanding." We sketch such a conceptualization in the following Section.

SECTION THREE

For the past 10 years, work has been underway at the Curriculum Laboratory of the University of Illinois concerned with observing student and adult behavior in relation to various mathematical tasks, and relating these observations to postulated knowledge representation structures and processing mechanisms. This Section reports on a few of these studies. More complete reports can be found in Davis (1982-A), Young (1982), Davis (1980-B), and Davis, Jockusch, and McKnight (1978). We shall also draw on work done elsewhere, which we identify at appropriate points.

VIII. Postulated Structures and Mechanisms.

The postulated structures and mechanisms can be divided into five categories:

- (i) Those concerned with representations for a specific problem, task, or situation;
- (ii) Those concerned with storage in memory and retrieval from memory;
- (iii) Those concerned with problem-solving in the sense of setting up a structure of goals and sub-goals;
- (iv) Algorithms;
- (v) Those concerned with making judgments about the correctness or usefulness of retrievals, goals, and representations.

We begin with a consideration of representations.

IX. Representations

A. In order to work on a mathematical task, one must represent the situation in some way. For very simple problems, representation may seem relatively unimportant, but as soon as a problem assumes even moderate complexity, representation of the problem situation becomes critical, and seems, for many students, to pose the most severe challenge to their ability to deal with the problem.

Consider this example, from a widely-used calculus text:

> A rope with a ring in one end is looped over two pegs in a horizontal line. The free end, after being passed through the ring, has a weight suspended from it, so that the rope hangs taut. If the rope slips freely over the pegs and through the ring, the weight will descend as far as possible. Assume that the length of the rope is at least four times as great as the distance between the pegs, and that the configuration of the rope is symmetric with respect to the line of the vertical part of the rope. (The symmetry assumption can be justified on the grounds that the rope and weight will take a rest position that minimizes the potential energy of the system.) Find the angle formed at the bottom of the loop.

Or consider this excerpt from a proof of the extreme-value theorem (grade eleven calculus in our Lab School):

> The proof involves the construction of a collection of closed intervals $I_1 = [a_1, b_1]$, $I_2 = [a_2, b_2]$, . . ., $I_n = [a_n, b_n]$, . . ., where each interval I_{n+1} is either the left or right half of I_n. The endpoints of these intervals are determined successively as follows: First of all, we take $a_1 = a$ and $b_1 = b$. To determine a_{n+1} and b_{n+1} in terms of a_n and b_n, we denote by c_n the mid-point of I_n [that is, $c_n = \frac{1}{2}(a_n + b_n)$] and we examine the function f over the two closed subintervals $[a_n, c_n]$ and $c_n, b_n]$. If there is a number x in $[a_n, c_n]$ such that $f(t) \leq f(x)$ for all t in $[c_n, b_n]$ we let $a_{n+1} = a_n$ and $b_{n+1} = c_n$. Otherwise we let $a_{n+1} = c_n$ and $b_{n+1} = b_n$. In the second case we note that for each t in $[a_n, c_n]$ there is at least one x in $[c_n, b_n]$ such that $f(t) < f(x)$.

In either case, you cannot begin serious work on the problem until after you have created at least a tentative representation for at least part of the problem situation, and a majority of the students whom we have observed have been brought to a halt at this point. (Note one important difference between the two problems: in the first problem, you know enough at the outset to make a drawing that must be reasonably close to correct; in the second, the information is so general that it does not determine, in detail, what the result will look like. Any sketch can, therefore, have only a "general" or "suggestive" character. A specific instance would resemble such a sketch in only a general way. It requires some sophistication to work with sketches of this type.) (Note 2.)

"Representations" are not merely something related to geometry. Any problem of more than trivial complexity requires some sort of representation, or at least one's performance on many tasks will be improved if one makes an appropriate representation. We have considered, earlier, the problem

$$\begin{array}{r} 7,002 \\ \underline{-\quad 25} \end{array}$$

and we shall consider, below, the problem:

Find f (f(x)), if

$$f(x) = \frac{1 - x}{1 + x} \quad .$$

Neither of these, as stated, is "geometric" -- but in both cases it turns out that certain geometric representations seem to be especially powerful.

To avoid misunderstandings, two further clarifications are needed:

1. A representation may be a combination of something written on paper, something existing in the form of physical objects, and a carefully constructed arrangement of ideas in one's mind. Our interest is especially aimed at ideas in one's mind, but some information will often be stored on paper, if only to reduce the strain on one's short-term memory. A situation on a chess board combines physical materials (the board and the chess pieces) with an elaborate idea in one's head (which, among other things, includes the rules of chess, and procedures for developing a strategic plan). Some players can, of course, play "mental chess," without

using a board. Every aspect of the board situation must then be carried by some sort of mental representation. Arithmetic commonly uses notations on paper, but some people can carry out mental arithmetic, again using only representations that they construct in their minds.

2. We are concerned with the representation of the situation or data for a single specific problem. But, of course, underlying the creation of such a representation is the general ability to build representations for certain situations. Thus, someone who knows about derivatives and integrals in calculus can easily construct representations where changes, rates of change, and changes in the rate of change are involved. Without such knowledge, constructing appropriate representations can be far more difficult, as one hears daily in political and economic discussions. For example, there are reports that "the rate of inflation is slowing down" -- this (usually) does not mean that prices are coming down, but rather that the percent increase in prices this month will be less than the percent increase in prices last month. (Question: Does this mean that the first derivative is positive, but the second derivative is negative?) In a similar way, there is ambiguity in the phrase "that clock is fast ". It may mean that the clock, capable of keeping correct time, has been set incorrectly, or it may mean that the clock is defective and the hour hand rotates through 360° in less than sixty minutes. This is an important ambiguity, which forcefully hits anyone who tries to make a correct representation for the statement. Because most people have not developed much capacity for creating representations involving rates of change, this ambiguity usually passes unnoticed.

As a second example, notice how much information can be conveyed by saying "this value exceeds the mean by more than two standard deviations" -- but only to a listener who possesses the ability to construct representations using such concepts as "distribution," "mean," "standard deviation" and so on!

A person playing "mental chess" must have a representation for the specific board position in this specific game at this specific moment -- and we shall focus attention on what can be said about such quite specific representations for the specific present problem -- but of course one is not unmindful of the underlying ability to construct such representations.

B. The function

$$f(x) = \frac{1 - x}{1 + x}$$

is weird, but its weirdness may not be immediately apparent. In fact, if

$$y = \frac{1 - x}{1 + x}, \quad (1)$$

one can easily solve equation (1) for x, and thus find the surprising result

$$x = \frac{1 - y}{1 + y} \quad .$$

This function is its own inverse! How can that be?

It may, at first, seem impossible. Most functions certainly do not behave like that. Consider $f(x) = 2x$. If you double a number, you cannot merely double it once more in order to get back to where you started -- e.g.,

$$3 \longrightarrow 6 \longrightarrow 12 \neq 3.$$

If you square a number, squaring once more won't get you back to where you started: e.g.,

$$5 \longrightarrow 25 \longrightarrow 625 \neq 5.$$

But for this weird function

$$y = \frac{1 - x}{1 + x},$$

this pattern does work: e.g.,

$$3 \rightarrow -\frac{1}{2} \longrightarrow 3 \qquad \text{Voila!}$$

Now ask: are there any other functions that behave this same way.

$$\left\{ \begin{array}{l} \sim \forall_x \; f(x) = x \\ \forall_x \; f(f(x)) = x \end{array} \right.$$

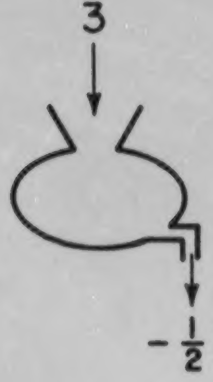
Even further: can you find a necessary and sufficient condition for a function to behave this way?

Nearly everyone who successfully deals with these questions necessarily makes use of one or more representations, of which the three most important seem to be:

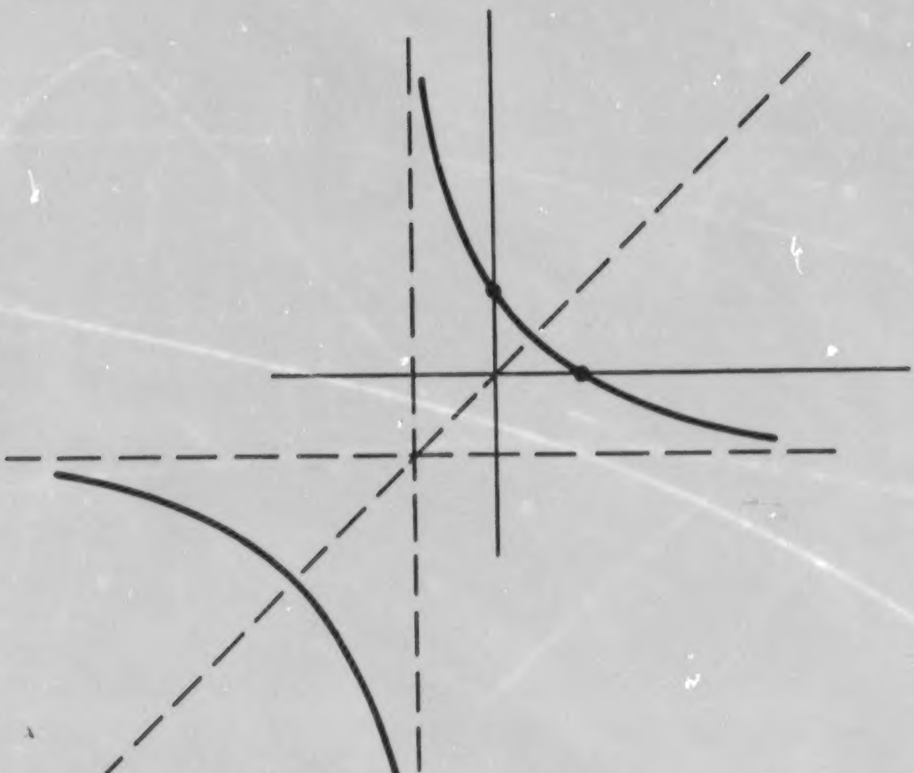
(a) "cycles":

$$3 \longrightarrow -\frac{1}{2} \longrightarrow 3, \quad \text{etc.}$$

(b) "machines" with "inputs" and "outputs":



(c) graphs (Cartesian coordinates)



There is an important aspect of an algebraic representation: If $f(x,y) = c$, then it is NAS that $\forall_x \forall_y \; f(x,y) = f(y,x)$. Thus

$$x^{1/3} + y^{1/3} = 2$$

qualifies, and so does

$$x + y + xy = 1$$

(which is another form of

$$y = \frac{1 - x}{1 + x}\,) .$$

Notice:

(i) The crucial role that the representations play in any analysis;

(ii) That certain general representation-creating capabilities -- such as the ability to graph functions -- can be crucial;

(iii) The prominent role played by certain "naive" or "pre-mathematical" representations (or metaphors), such as the "machine" with an input hopper and an output spigot.

X. The Special Role of "Pre-Mathematical Assimilation Paradigms."

One of the most persistent themes in ten years of research has been the prominent role played by certain metaphors that allow one to think of some mathematical situation as some sort of "simple" or "pre-mathematical" entity or situation. Thus, one can think of the numeral

7, 3 1 2

as an array of a few pieces of wood -- using MAB blocks -- namely:

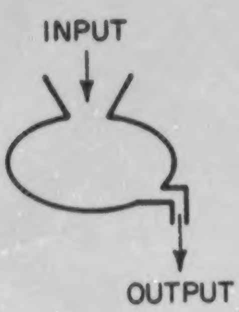
seven blocks
three flats
one long
two units.

A subtraction 7,002 - 25 can be thought of as an act of trading with these pieces of wood.

In our second example, the weird property

$$f(f(x)) = x$$

can be thought of in terms of the "machine"



These simple, frequently mechanical, metaphors are so prominent that we call them *pre-mathematical assimilation paradigms*.

Why are they so useful? Because they come equipped with a large collection of ancillary processing capabilities. As one example, consider *critics*. A *critic* (in this sense, as used by Herbert Simon and others) is an information processing procedure, stored in a person's memory, that becomes activated by certain information inputs, and responds by declaring that something has gone wrong. Thus, for most readers, the subtraction calculation

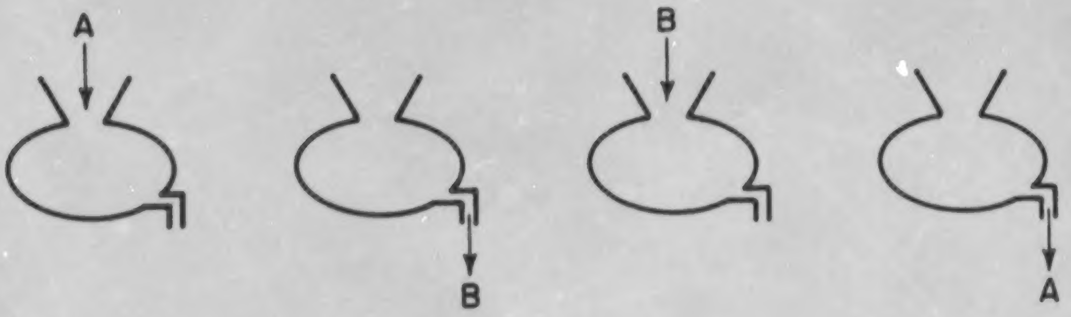
$$\begin{array}{r} 7,002 \\ -25 \\ \hline 5,087 \end{array}$$

will probably trigger a "size" critic, that will declare an error because "about seven thousand," minus twenty-five, should not be "about five thousand." [No third or fourth grader in our study showed any evidence of having this particular critic included in their repertoire!] A pre-mathematical assimilation paradigm -- such as the MAB block representations, or the "hopper-and spigot" machine picture -- will, because of its basic familiarity, have a large array of associated *critics*. MAB blocks are especially effective, because when one trades correctly -- e.g., one "flat" is traded for ten "longs" -- the trade *looks* correct and *feels* correct. Physical mass is conserved, and so is physical volume.

Critics, of course, are not the only cognitive mechanisms that are made available by simple "pre-mathematical" metaphors. The "hopper-spigot" metaphor allows us to interpret

$$f(f(x))$$

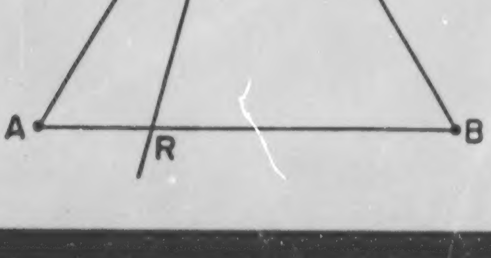
in terms such as this "cartoon-strip" representation:



Voila!

XI. When Primitive Paradigms Are NOT Available

The value of pre-mathematical assimilation paradigms becomes especially clear in those cases where no such paradigm is available. A particularly dramatic instance occurs when infinite sets of points are involved. One thinks of points as tiny specks of pepper or very little ball bearings -- or, to use one example we've encountered, bees. A long line segment, such as AB, contains exactly the same number of points as a very short segment, such as CD. One easily proves this with a few axioms ("Two points determine a unique line," and "Two lines, if not parallel, intersect in a unique point.") and with the diagram



which, properly interpreted, shows that every point R on AB corresponds to a unique point Q on CD, and vice versa. Students meeting this for the first time have difficulty, primarily because when one tries to think about points by using pre-mathematical metaphors more-or-less like ball bearings, the metaphor cannot be made to match the abstract situation, so incorrect cognitive apparatus is used again and again. Students try thinking about "smaller" and "larger" points, or try to "compress" the points more tightly together. None of these ideas is correct, and each must sooner or later be discarded.

Consider the spherical "ball"

$$x^2 + y^2 + z^2 < 1,$$

which is the interior space inside the spherical shell

$$x^2 + y^2 + z^2 = 1.$$

Compare this with bees (as one student did). Studies of bees show that they group into a large ball for protection against extreme cold. When the outer layer of bees become very cold, the bees shift positions; the cold bees move into the interior of the ball, and a new group of bees take their turn in the outside layer. Now, for $x^2 + y^2 + z^2 < 1$, which points (x, y, z) lie on the "outside"? Answer: none of them! Every point is protected by other points nearer to the boundary -- and these other points are themselves protected by still other points still nearer to the boundary -- and ... the process continues forever!

If such a configuration as

$$\{(x,y,z) \mid x^2 + y^2 + z^2 < 1\}$$

could exist as a physical entity, and if it were visible, and if points were opaque like tiny black marbles, what would you see if you looked at it? Although there were before your eyes a thoroughly dense packing of infinitely many points (uncountably many!), and even if every point could reflect light and was opaque, you could not see ANY points! Every point would be shielded from view by some other point that blocked your line of sight -- and you couldn't see the "blocking" points, either, because the view to them would be blocked by even nearer points . . . and so on!

XII. Young's Pyramids

A particularly interesting study of cognitive representations has been carried out by Stephen Young, inspired by a much-publicized wrong answer on an ETS test.

Consider, first, the general proposition. It frequently happens that one person will be able to answer correctly, and immediately, certain questions which most people could answer, if at all, only after considerable work. When the mathematician Hardy told Ramanujan that he had arrived in a taxi with a number plate that showed an uninteresting number, Ramanujan asked what that number was. Hardy said it was 1729. Ramanujan immediately answered that 1729 was not "uninteresting," since it is the smallest positive integer that can be expressed as a sum of two cubes in two different ways:

$$1729 = 1728 + 1 = 12^3 + 1^3$$

$$1729 = 1000 + 729 = 10^3 + 9^3 \quad .$$

How did Ramunujan arrive at this result?

Similarly, some people can answer immediately that

$$\frac{1}{\sqrt{2}}$$

is .707, that 8^3 is 512, and that

$$\int_0^\infty e^{-x^2}\,dx$$

is $\frac{1}{2}\sqrt{\pi}$.

In many of these cases, it turns out that the problem is easy if you represent it correctly. Thus

$$\frac{1}{\sqrt{2}} = \frac{\sqrt{2}}{2} = \frac{1.414}{2} = .707$$

$$8^3 = (2^3)^3 = 2^9 = \frac{2^{10}}{2} = \frac{1,024}{2} = 512$$

and

$$\int_0^\infty e^{-x^2}\,dx \int_0^\infty e^{-y^2}\,dy$$

$$= \int_0^\infty \int_0^\infty e^{-(x^2 + y^2)}\, dxdy = \int_0^{\pi/2} d\theta \int_0^\infty e^{-r^2}\, rdr$$

$$= \frac{\pi}{2}$$

which is the square of what was desired.

Norbert Wiener conjectured (and proved) that, if the function f(x) has an absolutely convergent Fourier series, and if $f \neq 0$, then

$$\frac{1}{f(x)}$$

has an absolutely convergent Fourier series. Why would anyone conjecture such a thing? It is NOT at all obvious -- when one thinks in terms of the usual representations! (How Wiener thought about the problem is not known.) But subsequent work by the Russian mathematician Gelfond on normed rings makes possible an alternative representation in which Wiener's theorem becomes immediately obvious.

Similar remarks could be made about Heaviside's "operator methods" in differential equations, and about many other parts of mathematics.

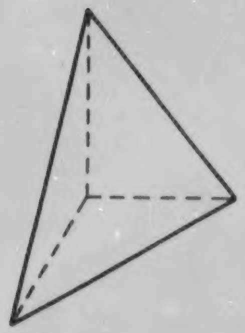
Now to the ETS problem . We are to construct two pyramids. Every line segment we will use has length L (they are all equal in length). Every face we will use -- with one important exception! -- is an equilateral triangle (every side has length L). The one exception is a square of side L. We now construct our two pyramids, one having this square base. If we now glue these two pyramids together by gluing together two congruent faces, how many faces will the resulting solid have?

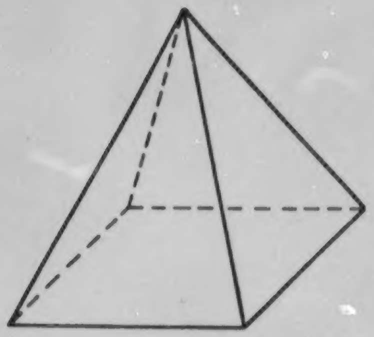
The ETS answer was 7, on the grounds that there had been 9 (4 on one pyramid, 5 on the other), and 2 had disappeared as a result of gluing.

One student said it was obvious that the correct answer was actually 5, because in two separate cases what had been two distinct faces would become single faces after gluing.

The question, of course, is whether a certain pair of two faces lie in the same plane or not. Do they?

All the experts to whom ETS had shown the problem considered it obvious that the faces did NOT lie in the same plane. If you think hard about the pyramids, you will probably agree. The angles don't work right -- the faces will not lie in the same plane.





But in fact the student is correct.

Young set himself the task of proving what might be called a "cognitive existence theorem" -- of showing that

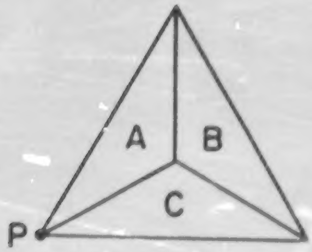
(i) There does exist a representation in which the correct answer is immediately obvious;

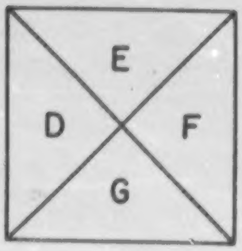
(ii) This representation is one which a student might reasonably build up from everyday experiences (if, of course, one has had the right "everyday experience" -- which, clearly, most people had not).

Here is Young's representation:

1. Most of us think about the pyramids with each pyramid sitting on a face, in a stable position. The physical stability itself guarantees that pyramids will usually be in this position when we see them.
2. If we think of the pyramids in this stable position, it seems unlikely that any exposed faces of one pyramid will be coplanar with any exposed faces of the other pyramid.
3. Indeed, one student gave this "proof" that they are not coplanar: "The key faces of the 5-sided pyramid are, in a sense, 'parallel'. The square base makes them move along, not getting closer to one another, nor further apart [as you move horizontally].

   "But for the all-triangle pyramid, the faces come together in a point. Therefore, the pairs of faces will not be coplanar." [The reader should make sure he or she sees which faces are involved in this discussion. Looking down from the top, the pyramids are



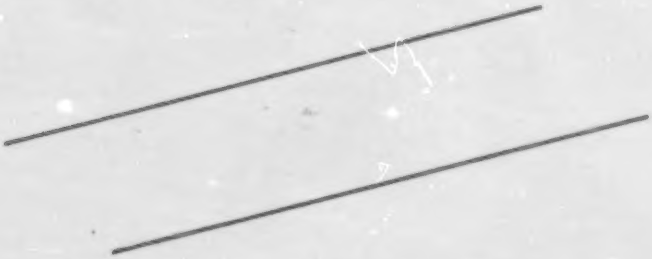


Suppose face B is glued to face D. The question, then, is this: Will face C and face G lie in the same plane, or not? (Similarly for face A and face E.) P is the point where the student in our study said the "faces [A and C] come together."

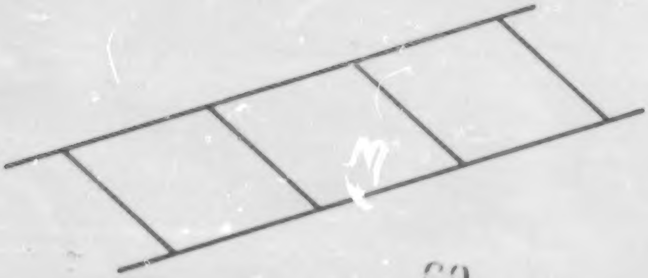
4. But tetrahedra do exist in other positions. In particular, some restaurants serve individual cream containers that are cardboard tetrahedra. One sometimes sees these piled up in disorderly arrays.
5. Without reproducing Young's complete analysis, here is his key alternative representation:

[The reader needs to try to see these perspective sketches as if they were really three dimensional.]

We have two parallel horizontal lines (shown in perspective):

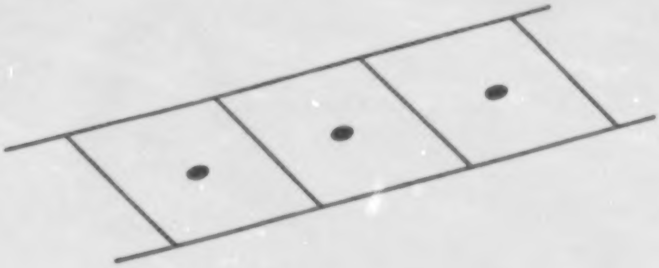


a row of squares:



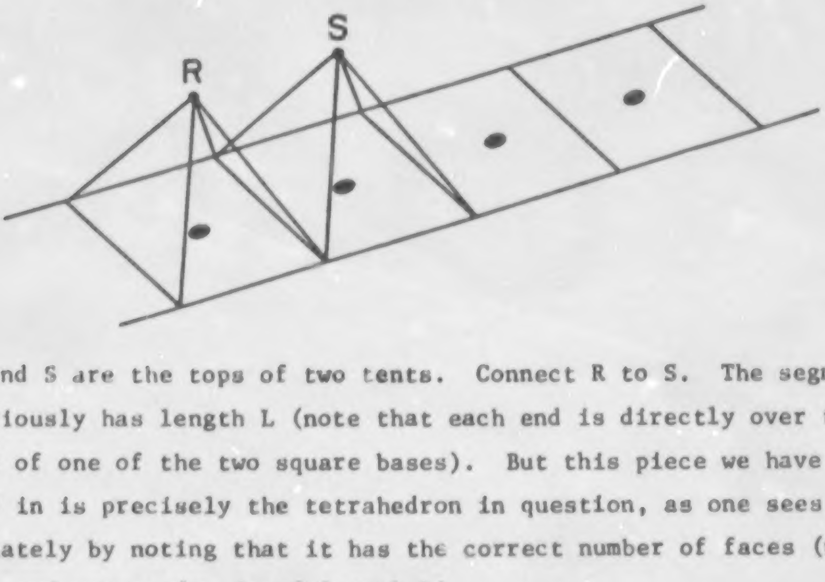
Each square is of length L along each edge (and, as before, every line segment we use will be of this same length L).

The midpoint of each square is located:



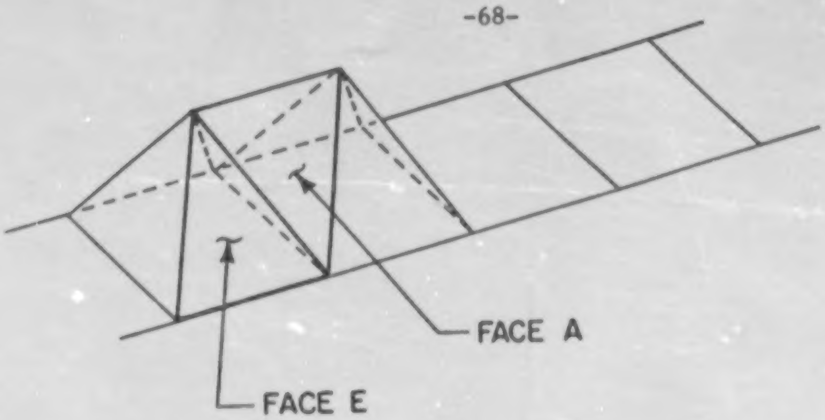
and at this midpoint a vertical "tent pole" is erected (this "tent pole" is the only exception to our requirement that each segment be of length L).

Using segments of length L, we now erect a row of "pup tents", each square being the base of one tent:



R and S are the tops of two tents. Connect R to S. The segment RS obviously has length L (note that each end is directly over the center of one of the two square bases). But this piece we have just filled in is precisely the tetrahedron in question, as one sees immediately by noting that it has the correct number of faces (namely, four), and every edge is of length L!

Now, in this representation, is there any doubt that the two triangular faces in question



-- face A and face E -- lie in the same plane? Of course not! Face A was constructed to lie in the same plane as face E! It has to!

What Young has accomplished is to prove that there does exist a representation for this problem situation in which the correct answer is immediately obvious. This representation might be called the "pup tents in a row" representation.

Notice that:

(i) This representation is simple enough that it can be constructed "in your head," without requiring the use of paper;

(ii) It is based upon pre-mathematical paradigms (which can be stated in terms of "tent poles," "pup tents," "single-serving cream containers," etc.);

(iii) It leads immediately to a correct solution.

The reader may feel that what is being presented in this paper is a weird version of mathematics. To be sure, this is NOT mathematics as most people know it. To the lay person who has not gone much beyond school arithmetic, mathematics consists of facts and algorithms that you learn by rote memorization. To the lay person who has had some success with high school algebra and geometry, a proof should consist of a two-column array of precise statements and axiomatic justifications. It should NOT deal with mental imagery about a row of pup-tents!

But notice that the Young representation immediately tells you how to construct a precise proof if one is called for!

Furthermore, there is a growing body of evidence that those who are "good at mathematics" do, in fact, think in terms of just the sort of "pictorial" metaphors that we have been discussing.

XIII. *Basic Conceptualizations*

The Greeks - perhaps the true parents of modern science - might have stared at islands and oceans and mountains for quite a long time, with little "scientific" result, had they not taken the decisive step of *postulating* some entities that made "space" more amenable to effective human thought: specifically, entities such as points, planes, lines, distances, etc.

A similar approach is a key to the alternative paradigm for the study of mathematical thought: the postulation of structures and procedures that represent mathematical knowledge and mathematical information processing. What should be postulated? Three dualities need to be considered: (i) the static representation of knowledge vs. the dynamic processing of information; (ii) "gestalt" or "aggregate" or "chunk" entities vs. sequential procedures; (iii) storage and retrieval of entitites, vs. the real-time *ad hoc* construction of representational entities. What each of these dualities means will become clearer in a few pages.

Presumably, before postulating things like "points" and "planes", the Greeks had thought quite a bit about physical space, and about possible intellectual tools that might make it easier to discuss space. In analyzing intellectual thought, before postulating comparable mathematical tools one needs to observe a large number of instances of human mathematical behavior. Such observations have been carried out (though vastly more are needed), - cf. e.g., Davis, Jockusch, and McKnight, 1978 - and the consideration of instances has suggested the postulation of several devices for processing mathematical information, which we now list.

A. *Sequential Processes*

1. *Procedures*. By "procedure" we mean an algorithmic, step-by-step activity, such as the cognitive sequence for adding 11 + 3 by starting with "eleven", then saying counting words to "count onward" from eleven by counting

the points of the symbol "3": "twelve" [→3]; "thirteen" [→3]; "fourteen" [→3]. "So eleven plus three is fourteen."

At least two kinds of procedures exist and produce observably different behavior.

(i) *Visually-moderated sequences* have the form of an input (usually visual) that cues the retrieval (from memory) of a procedure; execution of the procedure modifies the visual input; the modified visual input cues the retrieval of a new procedure; and the cycle continues until some process (possibly completing the solution) triggers termination. A very typical instance would be long division, being performed by someone who is not the full master of the algorithm:

A visual cue	$21\,\overline{)7329}$
triggers the retrieval of a procedure	["Uh, yes! How many times does 2 go into 7?"]
which produces a new visual cue (Note 3)	$\begin{array}{l} 3 \\ 21\,\overline{)7329} \end{array}$
which triggers retrieval from memory of another procedure	["Oh, yes! Now I multiply 3 times 21."]
which produces a new visual input	$\begin{array}{l} 3 \\ 21\,\overline{)7329} \\ 63 \end{array}$

. . . and the process continues.

Factoring quadratic polynomials is another example - again, if the student doing it is not yet the complete master of the topic:

The visual input	$x^2 - 5x + 6$
triggers the retrieval of a procedure	
that produces a new visual input	$(\qquad)\ (\qquad)$

which triggers the retrieval
of a procedure that produces

a new visual input (x) (x)

. . . and so on.

(ii) *Integrated Sequences*. With sufficient practice, a visually-moderated sequence can become independent of visual cues to trigger retrieval of the smaller component sequences which need to be strung together. Someone who knows long division well can describe the entire process without dependence on written cues. (They may, however, require paper as temporary storage for interim numerical results.) Sequences which, through sufficient practice, have become independent of visual cues for program guidance are called *integrated sequences*.

2. *Relations Among Procedures*. Within computer programming there is an important relationship among procedures: one procedure, A, may "call upon" or *transfer control to* a second procedure, B. When B has completed its assigned task, it returns control to procedure A. In such a relationship procedure B is said to be a *sub-procedure* of procedure A, and A is called the *super-procedure*.

It seems appropriate to postulate a similar relationship among procedures in the information processing that is part of human mathematical thinking. Indeed, observational data collected by Erlwanger (1973; 1974) indicate that, for the 6th grade students observed, errors were entirely in super-procedures calling for wrongly chosen sub-procedures. The sub-procedures themselves functioned correctly (cf. Davis, 1977). This, of course, is partly a comment on the school curriculum; over-learning of antecedent tasks had occurred satisfactorily, but the new tasks had not yet been mastered. Erlwanger's evidence from remedial tutoring suggested that, for most of these 6th graders, the 5th and 6th grade tasks probably never would be mastered.

As one example, one student answered .3 + .4 = ? with the answer

$$.3 + .4 = .07$$

The sub-procedure that accepted, as inputs, the numbers 3 and 4, and returned 7, operated correctly; so did the sub-procedure that counted "1 decimal place" and "1 decimal place" , and returned the format "2 decimal places" (i.e., .07). Of course, this second sub-procedure *should not have been called upon* in the process of solving the given addition problem; it should have been used only for multiplication.

In cases where a super-procedure called upon the wrong sub-procedure, Erlwanger's data showed a persistent relationship between the sub-procedure A (say) that *should* have been chosen, and the sub-procedure (call it B) that actually *was* chosen. Almost without exception, the visual stimuli to elicit retrieval of A and of B were extremely similar - for example, "3 + 3" vs. "3 × 3", or

$$\begin{array}{r} 10 \\ \underline{17} \end{array}$$

vs.

$$\frac{10}{17} \quad .$$

A further pattern has virtually no exceptions: within Erlwanger's data, it is very nearly always the case that sub-procedure B (which *was* chosen) is something that was learned *earlier* in the school curriculum. In other words, some recently-encountered *new* sub-procedure has erroneously been ignored, and its place has been taken by some more familiar 'old friend'.

An interesting explanation of this phenomenon can be given in terms of Minsky's theory of K-lines, but the details are complex, and quite beyond the scope of this article. (Cf. Minsky, 1980.)

B. *The General Problem of Flexibility.*

Thinking of offices, bureaucracies, and other human organizations, we all have some experience with the limits of flexibility within an organization; at some point, one reaches the boundary, and office procedure cannot deal appropriately with some specific instance because that instance lies outside of the original design for office procedures, and no further provisions for adaptation have been made.

Clearly, an analogous phenomenon bedevils human thought; one can reach a point where person A cannot cope, because no *explicit* procedures learned by A will suffice, and the creation of a new (and appropriate) procedure lies outside of A's capability.

To understand this phenomenon, most researchers attempt to postulate some definite body of procedures and knowledge representation structures (which we shall temporarily call "the system"), and then to distinguish "operations within a system" from "operations that involve stepping outside of the system". This distinction is made with exceptional clarity in Hofstadter, 1980. When a procedure orders up some sub-procedure, all of the activity is "within the system" (or "within the same level of the system"), rather as if a carpenter asks a fellow carpenter to hold a board in place while he nails it there. But clearly there are other kinds of operations that are needed.

A computer, asked to find the phone number of George Washington, first President of the United States, might call on a "Philadelphia" sub-procedure to scan the Philadelphia listings, or a "Virginia" sub-procedure to scan listings for Virginia, or even a "D.C." sub-procedure to scan the D.C. listings. That sort of thing could go on for a long time, unless there were some information-processing operators of a different type that were able to deal with different aspects of the task - for example, a "plausibility" operator that could make a historical check, perhaps calling on an "historical" subroutine that could query when Washington lived, and when the telephone was invented.

The original "phone directory" procedure, and its "Philadelphia" and "Virginia" sub-procedures, are on the same level (in this classification), a level that might be described as "finding the phone numbers". The "plausibility" operator, and its "historical" subroutine, are on a "higher" level (and on the same higher level), since they do not perform "phone-number-look-up" tasks, but "reflect" on the nature of such tasks. (To return to our carpenters, it is as if there were a "higher level" of operations, carried out by efficiency experts, architects, economists, etc., who do not cut boards and drive nails, but *who study the process* of cutting boards and building houses.)

This is an old issue in artificial intelligence and cognitive science; very often when a computer performs "stupidly" (as in spending vast resources in the search for George Washington's phone number), it is because the machine has been programmed with "task-performing" procedures, but *without* any higher level procedures to step back from, as it were, the assembly-line, and to look at what is going on.

There are various ways to provide for these "higher-level" operators, including at least these three (the first of which is NOT actually "higher-level"):

(i) checks may be inserted *at the original level* - e.g., before looking up any phone number. The look-up procedure can check dates, locations, and other reasons for believing that a phone number probably exists (this, of course, is not a real solution, because there remains the possible involvement of relevant attributes that have *not* been provided for in the checking procedure, as in the case of looking for the phone number of Hans Solo, or Lieutenant Uhuru, or KAOS);

(ii) procedures can be created that do not perform tasks on the original level, but exist on a higher level and *operate on* lower level procedures (a mechanism postulated by Skemp (1979) and by Hofstadter (op. cit.) and others.)

(iii) operations can take place in two separate areas: a "task-performance space" and a separate and distinct "planning space" (a solution implied by Simon, Minsky, Papert, and others).

All three solutions are possible for computers, although the second is, for computers, usually the most difficult; presumably all three are possible also for humans (who appear to make extensive use of the second method, which is one of the ways humans differ from today's computer programs.) It consequently seems desirable to postulate all three possibilities. We deal with the first method here (since it is really on the lowest, or "task-performing", level), and also postulate some mechanisms to provide for the second method. We defer discussion of the third method until later, when we deal with *heuristics* and *planning*.

1. *Critics.* A "critic" is an information-processing operator that is capable of detecting certain kinds of errors.

Example 1. A beginning calculus student wrote:

$$y = \sec x^2$$

$$dy = (\sec x^2)(\tan x^2)\ 2x$$

The teacher instantly recognized that there MUST be some error here, because a differential dy could not be equal to an expression which did not involve differentials. The teacher's collection of information processors included a *critic* which was not contained in the student's collection.

Example 2. We saw earlier Marcia's error in subtracting

$$\begin{array}{r} 7{,}002 \\ \underline{-25} \\ 5{,}087 \end{array}$$

Clearly, Marcia lacked "critics." By contrast, most adults possess a critic, related to the size of numbers, that should come into play here. After all, if I have about seven thousand dollars, and I spend twenty-five dollars, I should NOT end up with about five thousand dollars. Something is wrong! The third-grade student, however, believe her work to be correct.

One particular kind of information-processing arrangement, the so-called *production system*, provides an especially straight-forward method for dealing with *critics*. For a discussion of "production systems", refer to Newell and Simon (1972), or Davis and King (1977).

In both of the preceding examples we see differences in mathematical behavior that would be attributed to the presence, or absence, of certain specific critics.

2. *Operations on Procedures.* It is commonly postulated that a memory record is kept of procedures that have been used (Winograd, 1971; Davis, Jockusch, and McKnight, op. cit.; Minsky, 1980). It is also usually postulated that there are procedures which use the *sequence of active ("lower-level") procedures* as their *input*, and which *output* modifications of either *the collection of lower level procedures*, or else *the control structure*. John Seely Brown, for example, postulates a higher level operator that recognizes when the operational sequence is in a loop, and which intervenes in the control structure so as to terminate the loop. Other higher-level operators that have been postulated include a "look-ahead" operator that, with repetition, makes possible the prediction of which operator (or which input data) will be encountered next (Davis, Jockusch, and McKnight, op. cit.), and a "recognition" operator that can detect repetitions (idem). There is also evidence for a "simulate a run and observe" operator, as when a student, confronted with

$$x + 2\,\overline{\smash{)}\,x^4 + 2x^3 - 2x^2 - x + 6}$$

can "run through" in his head the algorithm for the long division of integers, as in

$$31\,\overline{)6851}\;,$$

"observe mentally" what happens, and thus solve the division-of-polynomials problem.

3. *Metaphor and Isomorphism.* Of course, underlying the ability to recognize the precise parallel between an algorithm for dividing integers and an algorithm for dividing polynomials, there is something far more fundamental: an ability to match up input data with some kind of knowledge representation structure that has been stored in memory. In typical information-processing explanations, four steps are postulated:

(a) Use of some cues to trigger the retrieval from memory of some specific knowledge representation structure;

(b) Mapping information from the specific present input into "slots" or "variables" that exist within the knowledge representation structure;

(c) Making some evaluative judgments on the suitability of the preceding two steps (and cycling back where necessary);

(d) If the judgement is that steps (a) and (b) have been successful, then the result is used for the next stage in the information processing.

One could illustrate these four steps as follows: if the task were to solve the equation

$$x^2 - 5x + 2 = 0,$$

then step (a) consists of observing some visual cues in the equation which cause us to say (in effect): "Aha! It's a quadratic equation!", with the result that we retrieve from memory the quadratic formula:

For the equation

$$ax^2 + bx + c = 0,$$

the solutions are given by

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

Step (b) now involves looking at our specific present input - namely, the equation $x^2 - 5x + 2 = 0$ - and taking from it certain specific information to enter into the "slots" or "variables" of our memorized rule. We see that "1" should be used as a replacement for the variable "a", that "-5" should be used as the replacement for the variable "b", and that "2" should be used as the replacement for the variable "c".

Step (c) involves whatever checks we carry out in order to convince ourselves that this is all correct, after which use of the quadratic formula (step (d)) easily produces the answers.

In one respect this example is too simple, and might therefore be misleading. In most examples of human information processing, the "knowledge" that is involved is more complex, and hence the knowledge representation structures are more complex. Suppose the content dealt, not with solving simple equations, but instead with reading and understanding a story about two people taking a trip of some sort:

> Leslie and Dana knew that they had several hours to travel, so they decided to seize the opportunity of having lunch.

The "variables" in this case deal not with simple numbers like 1, -5, and 2, but rather with such matters as: The sex of Leslie and of Dana (either could, after all, be either male or female); their mode of travel (horseback?

bicycles? flying in a commercial airliner? driving a car?); in what sense they are "seizing the opportunity" (stopping at an inn? stopping beside the road at a point where there is a good view? telling the stewardess that they do want lunch?); and so on.

Nor, by considering Leslie and Dana, have we left the domain of mathematics, which after all requires us to read words, sentences, paragraphs, equations, tables, and so on. In every case, however, the basic four-step operation is usually postulated as a fundamental part of the human information processing; in this, at least, information-processing theories do not treat reading and mathematics as being very different. The convenient accident that, within mathematics, the "slots" in knowledge representation structures *may* be labeled as mathematical variables "a", "b", "x", "y", and so on, is merely that: a convenience, but not an essential difference. (And, of course, a great deal of mathematical knowledge is stored in memory in other forms that do not make use of literal variables.)

One is dealing here with one of the most fundamental matters in human information processing. Presumably a successful selection, retrieval, and matching is what is meant by the word *recognize* - an instance of remarkable "folk" insight built into a common English verb. We have referred earlier to work by Hofstadter and by Minsky and Papert. Hofstadter (op. cit.) suggests this is what is commonly meant by "meaning", and Minsky and Papert (1972) remark that, when a situation on a chess board has been analyzed correctly, and (say) a "pin" has been recognized, it seems almost as if the pieces in question had suddenly changed color. The small pieces of input data are suddenly linked up with an important memorized data representation structure - the, small pieces have suddenly become a large "chunk" (in Miller's phrase). Instead of tiny "meaningless" bits of information, we now have a "chunk" to deal with, and it is this "chunk" which gives *meaning* to this aggregation.

These chunks are sometimes called "assimilation paradigms", and the teaching strategy that consists of *first* establishing such assimilation paradigms (or metaphors), then subsequently exploiting them, is sometimes called the *paradigm teaching strategy* (Davis, Jockusch, and McKnight, 1978). (Note that this educational use of the word "paradigm" is unrelated to Thomas Kuhn's historical use of this word!) This paradigm teaching strategy is used with stunning effectiveness by Hofstadter (op. cit.) to teach Gödel's theorem and the theorem that recursively enumerable sets are not necessarily recursive. For his "assimilation paradigm" metaphors, which he establishes beforehand, Hofstadter uses drawings and lithographs by M. C. Escher, and some "Lewis Carroll" style dialogues he created himself. The "paradigm teaching strategy" for pre-college mathematics is used in Davis (1980-A). In one example, a bag is partially filled with pebbles. By adding pebbles to the bag, by removing pebbles from the bag, and by interpreting the result as more pebbles, or less, than when one started, it is possible to give a meaningful interpretation of mathematical statements such as

$$4 - 5 = {}^{-}1$$

and

$$7 - 3 = {}^{+}4.$$

(Davis, 1967, pp. 57-61). Thus, the effect of the combined acts of putting 4 pebbles *into* the bag, and removing 5 pebbles *from* the bag, is to leave the bag holding one less pebble than it held beforehand. In the case of the second equation, the combined acts of putting 7 pebbles into the bag, and taking 3 out, produce the net effect of leaving the bag holding 4 *more* pebbles than it held before. The mental imagery of "putting pebbles into the bag" and "taking pebbles out of the bag" serves as a *paradigm* that guides the process of calculation; it is easy to demonstrate that this imagery serves this purpose considerably better than an explicit set of verbal rules can. (Notice that, if this "pebbles-in-the-bag" model is used, then the *knowledge representation structure* that is created within the student's memory is *not* like the quadratic formula, with explicit literal variables, but instead more closely resembles the kind of episode-based memory trace that one ordinarily associates with reading comprehension, rather than with mathematics.)

We turn now to knowledge representation structures in general.

C. *Knowledge Representation.* Several arguments establish the need to postulate, within knowledge representation structures, some form of *aggregates* or *chunks*. One of the most telling arguments is arithmetical: The number of possibilities that would need to be discriminated if, say, every symbol in a book were *independent* of all others, can be estimated as something like

$$70^{585000} = (70^{6})^{97500} \doteq 10^{1072500} ,$$

if one assumes 70 possible characters (a,b,c, ...; A,B,C, ...; 0,1,2,3, ...,9; plys punctuation and a "space"), 65 characters per line, 45 lines per page, and 200 pages. It is inconceivable that a human could discriminate $10^{1072500}$ different things; and one gets a hint of the difficulty if one imagines a 200-page book, where every symbol in every line on every page was a decimal digit, so that the entire book contained one huge number, 585,000 digits long. One could not "read" such a book because the human mind cannot process so much information.

Clearly, then, a book that contains 585,000 characters does *not* contain this much information. Rather, it contains far less because it is highly redundant. (As one trivial example, a symbol "q" in a word will necessarily be followed by a "u"; relatively few capital letters will appear; in the symbol string

space, t, h, ______, space

the blank can *only* contain "e" or "o", (with a few *very* rare exceptions).

But our assembling little bits of data into larger chunks goes much further than this; we normally deal in words, or sentences, or even larger aggregates - which, incidentally, is what makes proof-reading so difficult: we see what we are prepared to see, and this may not coincide with what is actually there.

1. *Frames.* But even more forceful arguments can be given that demonstrate that the information in one's mind must typically be organized into quite large aggregates (cf. Davis and McKnight, 1979; Minsky, 1975). For some of these larger aggregates, Minsky has used the word *frame* (although Rumelhart and Ortony use *schema*, and Schank uses *script*). A *frame*, then, is an abstract formal structure, stored in memory, that somehow encodes and represents a sizeable amount of knowledge.

A *frame* differs from a *procedure* in (among other things) the fact that a frame is not ordinarily sequential - it allows multiple points of entry, and provides some flexibility in its use.

2. *Retrieval and Matching.* Consider what needs to occur when an eleventh grade student is asked to solve the equation

$$e^{2t} + 6 = 5\,e^{t} \quad ,$$

or a calculus student is asked to integrate

$$\int \frac{e^{\sqrt{x}}}{\sqrt{x}}\,dx \quad .$$

The first student presumably knows how to solve

$$ax^2 + bx + c = 0,$$

and the second knows how to integrate

$$\int e^{u}\,du \quad ,$$

but each, of course, knows many *other* things that *might* be relevant to these tasks. In the quadratic equation case, a complicated process must take place, leading ultimately to the realization that

$$e^{2t} + 6 = 5\ e^{t}$$

matches *exactly* the pattern

$$ax^{2} + bx + c = 0$$

if you write

$$e^{2t} - 5\ e^{t} + 6 = 0\ ,$$

and make the correspondences

$$\begin{aligned} a &\leftrightarrow 1 \\ b &\leftrightarrow -5 \\ c &\leftrightarrow 6 \\ e^{t} &\leftrightarrow x \quad . \end{aligned}$$

This matching process succeeds only because the correspondence $e^{t} \leftrightarrow x$ necessarily implies the correspondence $e^{2t} \leftrightarrow x^{2}$, as a result of the addition law for exponents.

A similar analysis shows what must occur in the case of the integration problem.

How correct retrieval can occur - and often occur almost instantaneously! - is a considerable mystery. Minsky's recent "K-Lines" theory may provide an answer (Minsky, 1980), but we do not pursue the fundamental retrieval question further in this report. (Below, however, we shall consider some heuristics which students can learn that can improve student performance on retrieval problems of this general type.)

A "frame", then, is a formal data-representation structure that is stored in memory, hopefully to be retrieved when needed. This retrieval often occurs almost instantly. In a well-known experiment, Hinsley, Hayes, and Simon (1977) found that, for some subjects, *merely the first three words* ("A river steamer ...") *in the statement of a problem* were sufficient to trigger the retrieval of an appropriate frame (indicated by the subject interrupting after the third word, to say something like: "It's going to be one of those river things with upstream, downstream, and still water. You are going to compare times upstream and downstream - or, if the time is constant, it will be the distance").

Frames possess considerable internal organization. Especially important are the *variables* (or "slots") for which the frame will seek *specific values* from input data. When the present input does not provide enough information to permit certain slots to be filled, the frame may insert some tentative "guess", based on past experience. When slots are filled in this way, we say they contain "default evaluations" - data inserted (from past experience) to make up for gaps in the present input. Thus, in the earlier story about Leslie and Dana, if we knew they were on horseback, heading for a cattle drive in the old west, we might assume they were both male - but, of course, we could be wrong. Default evaluations are not guaranteed!

Default evaluations are important primarily in non-mathematical frames, where the matching of input data into slots is usually approximate; *it is a peculiarity of mathematics to require that every matching must be complete and precise!*

3. *Pointers*. Following computer practice, it is usually postulated that one mechanism by which one data representation structure can be related to another is the device of *pointers*; in effect, when a certain structure has been retrieved from memory and rendered active, and when certain definite conditions are met, then a *pointer* causes the retrieval of some other data representation structure, and/or causes some specific change in control. Pointers can also be used *within* a data representation structure to "weld" the whole unit together.

D. *Planning Language, Planning Space and Meta-Language*. It seems clear that complicated problems are often dealt with by carrying on two somewhat separate activities: actual calculations, and the process of *planning* what calculations are to be performed and how. To provide for this, it is common to postulate:

(i) descriptors that identify the possible uses of "action-level" processors;

(ii) mechanisms for dealing with descriptors (which can include tree searches, backward-chaining, etc.)

Example: Suppose a student encounters the problem: The plane P passes through the points A (a, 0, 0), B (0,b,0), and C (0,0,c). Find the distance from the origin (0,0,0) to the plane P.

Suppose also that this is, for the student, a *novel* problem, one that is *not* already familiar.

Presumably the student has the requisite knowledge to solve the problem, but this knowledge is scattered among the many techniques that the student knows. The task, then, is to select the correct techniques, and to relate them correctly to one another. Hopefully, the student possesses, for example, a technique that might be described as:

> How, given a non-zero vector V, one can find a unit vector $\vec{u}$ that is parallel to $\vec{V}$;

and another technique that might be described as:

> How, given the equation
>
> $$e\,x + f\,y + g\,z = k\ ,$$
>
> to find a vector that is perpendicular to the plane represented by this equation;

and so on. By sequencing these techniques correctly, the student can solve the problem, even though it is an entirely new problem that the student has never seen before. (For details, cf. Davis, Jockusch, and McKnight, op. cit.)

XIV. *Applications*

How can such conceptualizations, imported from cognitive science or artificial intelligence, be useful in studying the process of carrying out mathematical tasks, or learning to do so? We consider here some specific studies of this type.

A. *Some Specific Frames*

Can we identify some specific "frames" (i.e., knowledge representation structures) that most students build up and store in memory? The answer is: yes. From some general rules about how frames are created, one can deduce some probable frames; from the existence of certain frames, one can deduce observable behaviors. One can then check actual student performance protocols, to see if these behaviors do in fact occur.

1. *The Undifferentiated Addition Frame*. A common law of frame creation, used by Feigenbaum and others, is that discrimination procedures are no finer than they need to be. In the first year of elementary school, children typically learn (at least at first) only one arithmetic law, namely, addition. Hence they presumably synthesize a frame that will input 3 and 5 and output 8. When this frame is invoked, it will demand its two numerical inputs; it will, however, *ignore* the operation sign "+" because it has no need to consider this sign. There being only one arithmetic operation, discrimination among operations is not necessary.

When, in later months (or years), students encounter, say, 4 × 4, one should expect the wrong answer "8", and *this is by far the most common wrong answer* (cf., e.g., Davis, Jockusch and McKnight, op. cit.).

There is further evidence of the operation of this frame: Friend (1979) reports the seemingly curious fact that, of the three addition problems,

(A)	(B)	(C)
235	235	235
14	45	114
12	42	12 ,

problem (A) is the most difficult for elementary school students in Nicaragua, whereas, naively, one would expect (A) to be the easiest. After all, problem (B) involves one additional "carry" from one column to another, and problem (C) involves one more addition (the "2 + 1" in the left-most column).

In terms of frame operations, however, one would assume that a "column-addition frame", now being learned and tested, will make repeated sub-procedure calls on the primary-grade addition frame, *which demands two numerical inputs.* When, in problem (A), this demand is frustrated in the left-most column, a general law postulated by John Seely Brown predicts some intervention in the control sequence so that the program can be executed (see also Matz, 1980); this intervention distorts the column-addition pattern so as to obtain the required second input for the primary-grade addition frame, often by picking up a numeral from the "tens" column, so as to get the "wrong" answer 361 (taking "1" from the "12"). (For details, see Friend, op. cit.)

2. *The "Symmetric Subtraction" Frame.*

Consider another arithmetic operation learned in the primary grades: subtraction. At first, subtraction problems are of the form 5 - 3, but are *never* of the form 3 - 5. Hence, once again following the Law of Minimum Necessary Discriminations, students synthesize a frame that inputs the two numbers "3" and "5", and outputs "2". The frame ignores order since a consideration of order has never been important.

In later years, of course, the student will need to deal with both 7 - 3 and 3 - 7, and will need to discriminate between them. Such discrimination capability has not been built into the frame (which is why it is called "symmetric"). Consequently, in later years certain specific errors are easily predicted - and are, in fact, precisely what one observes (cf. Davis and McKnight, 1979).

(In a similar way, many adults are confused between "dividing into halves" and "dividing by one-half". The most common answer to 6 ÷ 1/2 is the wrong answer "3".)

3. *A Possible "Recipe" Frame.*

Karplus, in some elegant (and not-yet published) studies of ratio and proportion, reports protocols such as the following:

> In a story presented to the student, there is a boy, John, who is making lemonade with 3 teaspoonfuls of (sweet) sugar and 9 teaspoonfuls of (sour) lemon; a girl, Mary, is also making lemonade, but using 5 teaspoonfuls of (sweet) sugar and 13 teaspoonfuls of (sour) lemon. Appropriate illustrative pictures accompany the story. The interviewer asks: "Whose lemonade will be sweeter?"

1. Student: Let me see. Mary's would be sweeter.
2. Interviewer: Mary's would be sweeter? Um-hum [thoughtful tone invites further explanation . . .]
3. Student: Because Mary's has two less lemons in contrast with this [pointing to pictures], with John's.
4. Interviewer: Actually, she has 4 *more*.
5. Student: She *does?* [tone of great surprise]
6. Interviewer: [explaining his preceding remark] Well, she has 13 compared to 9.
7. Student: Yeah, but in relation to the *sugars*.

8. Interviewer: Could you explain to me how you figured that she has 2 less in relation to sugar?

9. Student: Well, O.K. . . . There's 3 and 9.

10. Interviewer: Uh-huh.

11. Student: And 3 goes into 9 3 times, and then you go 5 and 13 . . .

12. Interviewer: Yes . . .

13. Student: 15 goes into 5 3 times [sic!] , so it's really too much . . .

14. Interviewer: Uh-huh. So it's 2 less. And so if Mary wanted to make it come out the same sweetness as John's . . .

15. Student: It would have to be 15.

16. Interviewer: She'd have to use 15. So, I see she has 2 less. O.K.

This interview excerpt can be split into three sections: utterances 1 through 7 show frame-like behavior, utterances 8 - 13 show sequential behavior under frame control, and utterances 14-16 are the interviewer's attempt to re-state the student's idea in more explicit language.

The student's comparison of the table

	Sugar	Lemon
John	3	9
Mary	5	13

with a different table *which is nowhere in evidence except in her own imagination* is stunning! The alternative table, namely,

	Sugar	Lemon
John	3	9
Mary	5	15

is so real to her that she at first rejects the interviewer's "common sense" numbers - *which ARE the numbers that are actually in evidence* - in favor of her "ideal" table.

This, and other evidence in various Karplus interviews, suggests strongly that many students have a "recipe" frame, which allows them great facility in doubling recipes, halving recipes, etc.

4. *A "Units" or "Labels" Frame.*

Clement, Kaput, and their colleagues (Clement and Kaput, 1979; Clement and Rosnick, 1980; Lochhead, 1980) have carried out an important series of experiments dealing with student responses to this question: "At a certain university, there are six times as many students as there are professors. Please write an algebraic version of this statement, using S for the number of students and P for the number of professors."

The correct answer, of course, is 6P = S; but an exceedingly common wrong answer, even among engineering students, is 6S = P.

By itself this might mean very little. After all, humans are fallible, and 6S = P is almost the only wrong answer that any reasonable person would invent. Furthermore, one could explain the error as a manifestation of internal information processing that simply follows the time-sequential order (or left-right order) "there are SIX times as many STUDENTS as there are PROFESSORS" (or a possible abbreviated version: "SIX STUDENTS for each PROFESSOR").

But the phenomenon is far deeper than this: Clement and his co-workers have varied word order, studied different populations, used different degrees of "meaningfulness" (after all, there are almost always more students than professors - but how about the ratio of sheep to cows in a certain farm?), and varied the ratio (e.g., "there are five professors for every two students"). Most importantly they have tape-recorded interviews in which a tutor attempted to correct the error in students who were writing the wrong equation. No simple "accidental slip" is involved here, as one sees from four facts:

(i) Students who are initially wrong protest vigorously against the change. ("I can't think about it that way!" "You're getting me all mixed up!" "That's weird!");

(ii) Students who are initially wrong are very reluctant to change, and if they do change to writing correct equations, they soon slip back to the wrong versions;

(iii) In order to preserve their wrong equations, students would make egregious variations in the definition of variables, even concluding that "S" must stand for the number of professors (sic!) and "P" must stand for the number of students (or "S" must stand for the number of cows, and "C" must stand for the number of sheep);

(iv) Students who wrote the wrong equations tended to *verbalize* the problem differently from students who wrote the correct equation: students writing the correct equation tended to say "Six times *the number of* professors equals *the number of* students", whereas students writing the wrong equation tended to say "There are six students for each professor."

In any situation of very common and very persistent errors of this type, one who believes in frames will quickly suspect the presence of a frame that is itself perfectly useful (and in that sense "correct"), but which is being retrieved when it should not be, and being put to use for a task where it is not appropriate. The fourth characterization above gives a strong confirmation of this conjecture, and even indicates what this alternative frame probably is: it is a frame that has been developed for dealing with *units* and with *labels*. We have all seen "equations" such as:

12 inches = 1 foot
3 feet = 1 yard
5280 feet = 1 mile
2.54 cm. = 1 in.

and so on. The labels "inches", "foot", "in.", etc., are *not* variables, and do not behave like variables. For contrast, let I be Mohammed Ali's height in *inches*, and let F be his height in *feet*. What equation can you write between I and F? (Cf. Davis, 1980-B.)

B. *The "Greater Than" Relation.*

Richard J. Shumway, of Ohio State University, has pointed out the discrepancy between formal definitions of

$$a < b ,$$

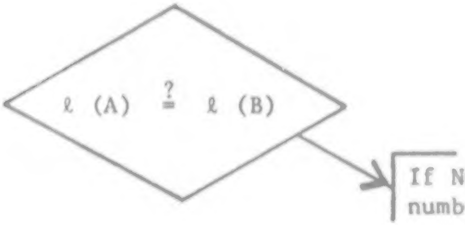
vs. what students actually do to decide whether $a < b$.

For example, one may define: "$a < b$ if and only if there exists a positive number N such that

$$a + N = b ."$$

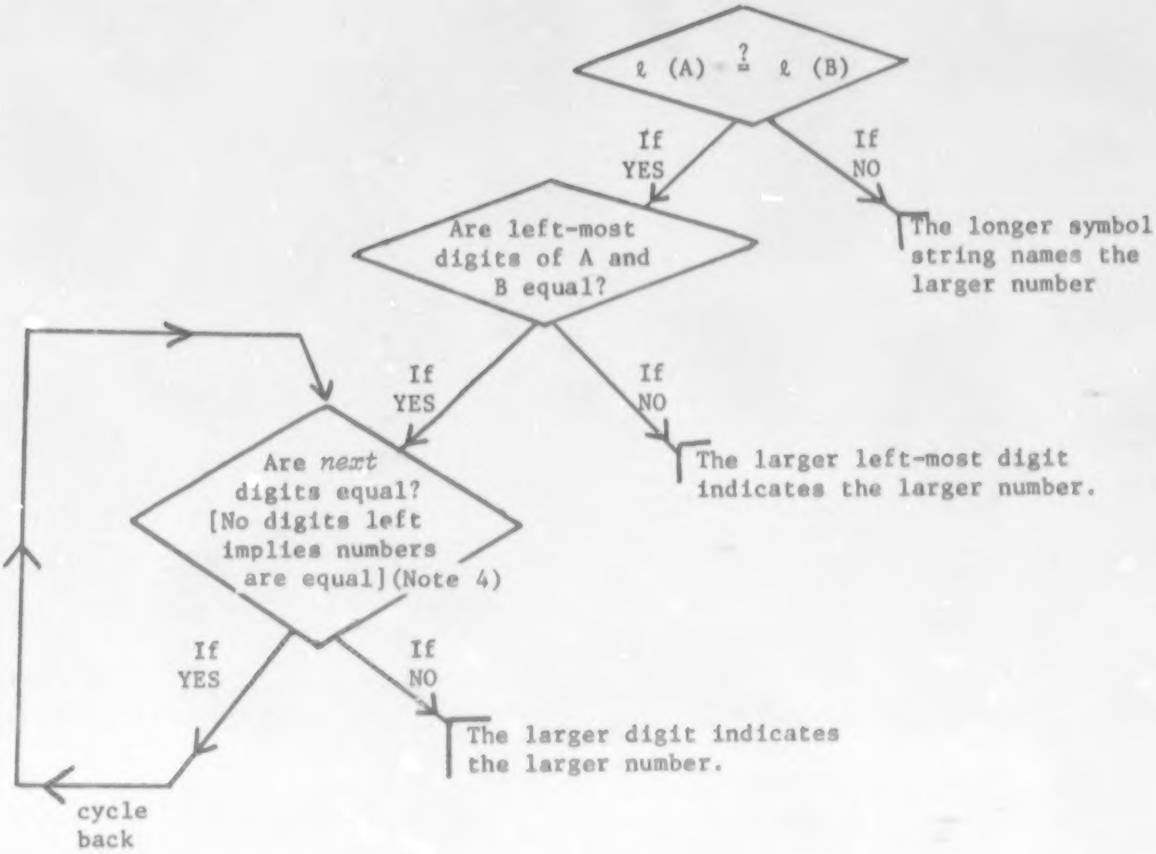
Now, ask a student which is smaller, 31 or 2,986. What thought processes will the student employ? Does he start adding numbers to 2,986 to see if he can get 31 as a result? If he did start such a search, how would he know when to give up?

a) For *positive integers*, the first step in deciding probably looks like this (where ℓ (A) denotes the number of digits in the decimal representation of the integer A, etc.)



If NO, longer symbol string names larger number.

b) But suppose ℓ (A) = ℓ (B); what then? Typical students presumably use some step-wise decision procedure - or else a production system - *roughly* equivalent to this:


ℓ (A) ?= ℓ (B)
If YES
If NO
The longer symbol string names the larger number
Are left-most digits of A and B equal?
If YES
If NO
The larger left-most digit indicates the larger number.
Are *next* digits equal? [No digits left implies numbers are equal] (Note 4)
If YES
If NO
The larger digit indicates the larger number.
cycle back

This is all very well for positive integers, but something more is needed to deal with negative numbers, rational numbers, etc. For example:

(i) Which is larger, -1,239 or -37?

(ii) Which is larger, 0.039 or 1.15?

(iii) Which in larger, 0.0395 or 0.00953?

We do not pursue this example further in this chapter. The main point of Shumway's example is that older analyses in terms of formal definitions or, cognitively, "having a concept", although these may be valuable for certain purposes, *are NOT as useful for understanding a student's thinking as the "procedure-and-frame" kind of analysis is. Clearly, a student does not find a positive number to add to 0.039 so as to get 1.15, in order to answer the* *question of whether 1.15 is larger than 0.039.*

C. *Deeper-Level Procedures*

It is a well-known fact (cf. Davis, Jockusch, and McKnight, op. cit.) that students who have learned to solve quadratic equations by factoring

$$x^2 - 5x + 6 = 0$$

$$(x - 3)(x - 2) = 0;$$

so, either $x - 3 = 0$,

or else $x - 2 = 0$;

hence, either $x = 3$ or else $x = 2$

tend to make the following mistake:

$$x^2 - 10x + 21 = 12$$

$$(x - 7)\ (x - 3) = 12;$$

so either $x - 7 = 12$ or $x - 3 = 13$;

therefore, either $x = 19$ or else $x = 15$.

This error is very difficult to eradicate - or, at least, very difficult to eradicate *permanently*. Even when classes of able students, using a seemingly excellent textbook, receive careful instruction - with emphasis on the special role of zero in the "zero product principle" - it is still the case that this error will continue to crop up in student work. Despite careful explanations of why it is an error, despite short-term elimination of the error, it keeps coming back.

Matz (1980) presents a theory of cognitive processing that explains the persistence of this error. She postulates two levels of procedures (stateable as "rules"). The surface level rules are ordinary rules of algebra; the deeper level rules serve the purpose of *creating* superficial-level rules, *modifying* superficial-level rules, or *changing the control structure*.

Now one of the deeper level rules must surely generalize over number; that is, it must say, in effect,

$$[P(a), P(b), P(c)] \Rightarrow [\forall_x P(x)].$$

In other words, if I show you how to add

$$\begin{array}{r} 23 \\ +14 \\ \hline 37 \end{array} ,$$

you will never master arithmetic if you persist in believing that this works *only* for 23 and 14. You must believe that this same procedure works also for

$$\begin{array}{r} 34 \\ +15 \\ \hline \end{array}$$

or

$$\begin{array}{r} 34 \\ +25 \\ \hline \end{array} ,$$

and so on. In short, in order to learn arithmetic you *must* possess a deeper level rule of the type that Matz postulates.

Now, it is a known property of rules that, in their less mature forms, they tend to be applied too widely; the appropriate constraints on their applicability have not yet become attached to the rules.

The *zero product principle,*

$$[A \cdot B = 0] \Rightarrow [(A = 0) \vee (B = 0)], \qquad (1)$$

is very nearly the first law students have encountered where some specific number (zero) must not be generalized. Predictably, then, the lower-level "generalizing" rule will be used to extend equation (1) to read

$$[A \cdot B = C] \Rightarrow [(A = C) \vee (B = C)] \qquad (2)$$

Equation (2) would be a correct generalization of equation (1) *if generalizing were appropriate in this case.* Unfortunately it is not.

Then this error is hard to eradicate for the same reason that dandelions are - something is creating "new" dandelions, so even when you eliminate dandelions, the job is not permanently accomplished. Even when you achieve enough remediation that a student ceases - temporarily! - to commit the error, your work may not be finished, for the student has, in his own mind, the deeper level rules that are capable of creating anew the incorrect factoring pattern, in accordance with commonly-accepted laws of cognitive processes.

John Seely Brown has used this same notion of "deeper level rules" to explain the observed fact that multiple errors occur in arithmetic problems more often than ordinary probabilities would predict. Brown postulates that a first "bug" (i.e., systematic error) may produce a control error (such as an infinite loop); an "observer" procedure notes this control error, and intervenes; the intervention takes the form of modifying the control procedure so as to exit from the infinite loop; but this intervention will itself tend to create further performance errors. Consequently, multiple errors occur disproportionately often in student work.

D. *Knowledge in Other Forms; The Semantic Meaning of Symbols*

One who is skillful in the performance of mathematical tasks must be able to work with mathematical symbols in a relatively "meaningless" way, guided only by patterns and formal rules, but they must also be able to deal with the meanings of the symbols - or at least, with some of the meanings of the symbols. We have seen earlier that third grader Marcia would subtract

$$\begin{array}{cccc} 7, & 0 & 0 & 2 \\ & - & 2 & 5 \\ \hline \end{array}$$

incorrectly, following the standard algorithm, but using a version of that standard algorithm that contains a "bug":

$$\begin{array}{cccc} 6 & & & \\ \cancel{7}, & 0 & 0 & {}^{1}2 \\ & - & 2 & 5 \\ \hline & & & 7 \end{array} \quad ,$$

$$\begin{array}{cccc} 5 & & & \\ \cancel{6} & & & \\ \cancel{7}, & 0 & {}^{1}0 & {}^{1}2 \\ & - & 2 & 5 \\ \hline 5, & 0 & 8 & 7 \end{array}$$

In a recent study, Davis and McKnight (1980) used this subtraction problem as an interview task in order to compare children's algorithmic knowledge (which could, of course, be rote), with the knowledge which these same children had concerning the meaning of the various symbols and operational steps.

The interviews sought to study each child's possession of five kinds of knowledge that might be labelled "meaningful". The interest in this question arises from a question of knowledge representation; the different kinds of knowledge would require coding into different representational forms.

The five forms of knowledge were as follows:

i) The *size* of the numbers should immediately signal an error - "about seven thousand" minus "a few" ought not to turn out to be "about five thousand". This would be coded in "critic" form. Interview data showed that no third grader interviewed (from four different schools) had this level of understanding of approximate *sizes* of numbers. Since the students did not possess this kind of understanding, it could not be used to help identify and correct the error in the algorithm.

ii) *The use of simpler numbers.* Perhaps numbers like "seven thousand" are essentially meaningless to these students; perhaps, then, smaller numbers would be more meaningful, and might thus allow *meaning* to guide the algorithm. This kind of knowledge would be coded as a heuristic strategy. Unfortunately, since the error in question involves "jumping borrowed one's over zero's", it is not possible to use truly small and familiar numbers - but one can, at least, use *smaller* numbers. For example,

$$\begin{array}{r} 702 \\ \underline{-\ \ 25} \end{array}$$

No improvement in student performance was achieved by switching to smaller numbers.

iii) Adults who regularly use "mental arithmetic" to solve such problems *without writing* do not usually do so by visualizing the standard algorithm. On the contrary, they take advantage of the special properties of specific numbers. E.g., for 7,002 - 28, one can say: 7,000 - 25 would be 6,975. But if I subtract 3 more (because 28 = 25 + 3), I will end up with three less, so 7,000 - 28 must be 6,972. But if I now *start with two more*, I must end up with two more, so 7,002 - 28 must be 6,974. This kind of knowledge would be procedural. In interviews, this general method was taught to students. Results: (a) Those students who learned it well enough to get *correct* answers to 7,002 - 28, etc., nonetheless had more confidence in the correctness of their (wrong) algorithmic answers; (b) students more often attempted to visualize the usual algorithm - in short, they persisted in unchanged algorithmic behavior, with only the modification that they attempted to visualize the algorithm instead of actually writing it down on paper; (c) in no case was this kind of knowledge used to correct the "bug" in the algorithm.

(iv) "Borrowing", as in

$$\begin{array}{r} \overset{6}{\cancel{7}},{}^{1}5\ 9\ 2 \\ -\ \ 6\ 2\ 1 \\ \hline 6,\ 9\ 7\ 1 \end{array}$$

can be interpreted as "giving the cashier a thousand-dollar bill, and receiving in exchange 10 hundred-dollar bills". This kind of knowledge would be coded as a frame based on past experience. The interviews revealed that every third grader could deal correctly with such "cashier transactions" when they were presented directly (and not by implication, in a subtraction problem) - that is, if asked "How many hundred-dollar bills could you get in exchange for a thousand-dollar bill?", every third grader answered such questions correctly. However, no third grader saw the relevance of this to the subtraction algorithm, and none sought to correct the algorithmic error as a result of the discussion of "cashier exchanges". (The interviewers carefully *avoided* suggesting how the two might be related, since it was the goal of the study to see if the students would *spontaneously* see the relevance of "cashier exchanges" to the subtraction algorithm.)

v) *Dienes' Multi-base Arithmetic Blocks.*

Dienes' MAB blocks provide a physical embodiment for place-value numerals, and allow a physical "subtraction" that is precisely analogous to the subtraction algorithms (cf. Davis and McKnight, 1980). This would also be coded as a frame. In one school included in this study, students did learn how to represent 7,002 correctly in terms of MAB blocks ("7 blocks and 2 units"), and similarly for 28 ("2 longs and 8 units"), and could interpret the task 7,002 - 28 in MAB terms ("you have 7 blocks and 2 units, and you are asked to give someone 2 longs and 8 units"). Nonetheless, no third grader saw (i) that this indicated an error in their algorithmic calculation, or (ii) that this MAB task showed how to *correct* the error in the algorithmic calculation.

The over-all result was that the students were entirely wedded to an algorithmic performance of this subtraction task, preferring their algorithmic answers (even when wrong) to answers obtained in other ways (even when these answers were in fact correct). Asked to check their answers, they merely repeated the algorithm. No knowledge of possible *meanings* of the symbols was brought to bear on the algorithmic task. *None!*

It is interesting to compare this result to a persistent theme that emerges from many studies by Ginsburg, who reports that students commonly possess important mathematical competences of "non-school" origin which they do not relate to school tasks. (Cf. Ginsburg, 1977; 1981-B.)

Implications for Teaching. Does this mean that, say, MAB blocks are no help in learning algorithms? No, surely not. Many experienced teachers believe that the MAB blocks can be quite helpful. Furthermore, in general, mathematically-experienced people frequently report being guided in their calculations by a knowledge of the meaning of the symbols. Presumably what this unexpected outcome does indicate is that the school in question was not relating MAB blocks to algorithmic calculations, so that, even though the students were getting a good knowledge of how to set up MAB representations, they were not *using* these representations to guide them through the algorithm.

Adults can easily see the students' point of view here: anyone who is following a very unfamiliar and complicated recipe for the first time may find themselves checking up primarily by checking through the recipe, one line at a time, to see if it seems to have been followed correctly - and this is exactly what the students did. One doesn't "think about the task in other terms" because one lacks the tools to be able to do so. With experience, all of this can change, and knowledge of other sorts can be brought to bear on the task at hand.

How Do "Other Meanings" Relate to the Theory? The study of third graders has been presented primarily in terms of observable behaviors. How would these phenomena be formulated in theoretical terms?

The algorithmic performance is the easiest to conceptualize - a programmable hand-held calculator can exhibit this kind of performance, and it is readily conceptualized as a "procedure" consisting of a sequence of simple "unit" steps.

The knowledge of MAB blocks which the students demonstrated is probably contained in a collection of procedures, any one of which performs some specific task - such as trading ten units for one long, or recognizing that

"7,396" calls for three flats (and also 7 blocks, 9 longs, and 6 units). (Of course, this could be in the form of relatively powerful, relatively general procedures that can deal with, say, any "10-for-1" trading situation, or it could be in the form of a larger number of more specific procedures, such as "trading ten longs for one flat".) Because such behavior shows little of the sequential rigidity of a procedure, it would be classified as "frame-like".

But for the students who had both the "algorithm" procedures and the "MAB" frame, and who nonetheless failed to relate the two, what was lacking?

There are several possibilities, including at least these:

i) The students may lack *pointers* in the algorithm procedure that would invoke the MAB block frame;

ii) The students may lack a goal-oriented control mechanism that would relate to the MAB blocks frame to establish goals for block exchanges, and to sequence exchanges so as to achieve these goals;

iii) The students may never have *reflected* on the MAB frame, and to discover pattern similarities between the MAB frame and the algorithmic procedures.

Concepts.

In the 1950's it seemed that mathematicians meant one thing by the word "concept", while psychologists and educators meant something else. Mathematicians spoke of "the concept of *function*" (Note 5) or "the concept of *limit*." By contrast psychological studies of "concepts" seemed to deal only with rules for inclusion in a certain class of things. (To be sure, it may appear that anything whatsoever *can* be formulated as a class inclusion problem, but this often distorts the reality so badly as to be positively harmful.)

Artificial intelligence, cognitive science, or even what is nowadays called "knowledge engineering", provides a way to express something that is closer to the mathematician's notion of a "concept". If you have mastered, say, the concept of "limit of an infinite sequence", you possess adequate knowledge representation structures of certain specific types, and you possess an adequate array of pointers to guide certain appropriate associations. You also possess a collection of useful examples (or the means of creating new examples), and an ability to relate examples to general statements.

For the concept of "limit of an infinite sequence", you would need at least knowledge structures representing:

i) the graph of $u_1, u_2, u_3, \ldots$ showing u_n vs. n, with a horizontal line for the Limit L, a strip representing $L - \varepsilon < u_n < L + \varepsilon$, and a representation of the "cut point" N (for $n > N$);

ii) the interpretation of

$$|u_n - L| < \varepsilon$$

in terms of the *distance* from u_n to L;

iii) an ability to convert between

$$|u_n - L| < \varepsilon$$

and

$$L - \varepsilon < u_n < L + \varepsilon$$

and the graphical representation (as in (ii), above);

iv) "metaphoric" language, describing ε as an "allowed tolerance", and N as a "cut point";

v) knowledge of the consequences of choosing ε first, and N second, vs. choosing N first, and ε second;

vi) metaphoric language to describe (v) intuitively;

vii) even more intuitive formulations, such as: "L is the limit of the sequence u_1, u_2, . . . if every term u_n in the sequence is equal to L - except that when I say "equal" I will forgive an "allowed error" ε, and when I say "every term" I mean "except for a finite number at the beginning";

viii) the usual ε, N definition;

ix) an ability to relate all of the preceding structures;

x) knowledge of how "*arbitrarily* close" works, or can be used;

xi) knowledge of how *indirect proofs* are employed, especially by using (x) above;

xii) knowledge of how (xi) uses the Law of Trichotomy;

xiii) the axiom or theorem that every Cauchy sequence converges;

xiv) a classification of sequences as: monotonic, increasing, non-decreasing, oscillating, convergent, divergent, bounded, and unbounded;

xv) either a collection of sequences that are examples of the categories in (xiv), or else an ability to generate examples as needed;

xvi) knowledge of various common errors, and precisely *why* they are errors. (E.g., the error of claiming that "the sequence 1,1,1, . . . is not convergent because the terms are not getting nearer to any number"; the error of claiming that "the limit of .9, .99, .999, . . . must really be less than one, because the terms of the sequence are always less than one"; the inadequacy of defining limit by saying "the limit is the number that the terms are getting nearer to"; the error in defining limit by saying "given any $\varepsilon > 0$, there must be an integer N such that there exists a term u_n, with $n > N$, and with $|u_n - L| < \varepsilon$"; the error in assuming that "the limit of a sequence is either an upper bound for the terms of the sequence, or else a lower bound.")

The relation between *general statements* and *examples* is so important that it deserves special attention (cf. Rissland, 1978,A, B). Student errors reveal something of this relationship. One 12th grade calculus student defined the *limit of a sequence* by writing:

> The limit of a sequence is a number that the terms approach but never reach.

This student had seen, in class, sequences such as

.9, .99, .999, . . .

and

1, 1.4, 1.4., 1.414, . . .

Even for these sequences, the student's answer is inadequate, but it fails flagrantly for sequences such as

1, 1, 1, . . .

or

1, 0, 1/2, 0, 1/3, 0, 1/4, ...

Subsequent interviews showed, unsurprisingly, that the student was not bringing to mind examples of this type to test the suitability of his definition.

The other main error in the student's answer can also be revealed by testing his statement against appropriate "test case" examples. The first error, which we have seen, involved his use of the phrase "but never reach", and is revealed by considering examples of possible *sequences*. The second error is revealed by testing his definition against examples of possible *limits*. Consider the sequence

.9, .99, .999, ...

Clearly, the terms of this sequence "approach" - that is, get nearer to - the number 1, which is what the student had in mind. But the terms also get nearer to the number 1.01, they get nearer to the number 1.5, they get nearer to the number 2, and, in fact, they get nearer to the number one million. To be sure, they never get very near to one million, but .999 is closer to one million than .9 is !

The relation between known examples and general statements is so important in mathematics (Rissland, 1978,A,B) that it seems necessary to postulate two information-processing capabilities:

1. Given a general statement, one can retrieve from memory, or construct, examples by which to test the statement. (The postulate does not assert how successfully this will be done in any particular case, only that in principle it is something that *can* be done - just as humans possess, in general, the ability to move from one place to another, whereas most plants do not.)

2. Given a collection of examples stored in memory, one can make general statements that describe common attributes of these examples.

Mathematical thinking is heavily dependent upon these two capabilities.

When a physicist or mathematician says "I must educate my intuition" about certain matters, it seems likely that part (at least) of this process means synthesizing data representations of appropriate examples, and establishing relations among them - as, for instance, when the phenomenon of a returning space capsule hitting the earth's atmosphere can be better understood by relating it to skipping flat stones across the surface of a lake. Both are useful examples - one familiar, and the other not - of the surprising ability of a solid to glance off of a liquid if its velocity is adjusted in a certain way.

XV. *Planning Space.*

We have considered earlier the problem: plane P passes through the three points A(a,0,0), B(0,b,0), and C(0,0,c). If a, b, c are all non-zero, find the distance from the plane P to the origin O(0,0,0).

This problem involves *planning* only if, as we suppose to be the case, it is a novel problem which the student has not previously encountered. As a novel problem, it is rather difficult. We further suppose, of course, that the student has learned the separate techniques (in vector form) needed for a solution. Thus, the student's real task is to select (and retrieve from memory) the correct techniques, sequence them correctly, and establish the proper relations between them.

This task is fairly easy, however, if and only if the student has the correct descriptors attached to each technique, and has developed the procedures needed for searching among these descriptors. One can think of this, informally, as if each procedure is a specific tool, and attached to each tool is a tag that describes what the tool can be used for. Execution, of course, requires the use of the tools themselves, but *planning* is carried out merely by reading the "tags" or "labels". One tag, for example, says: *if you have a vector V (of any non-zero length), and a unit vector* $\vec{u}$*, you can find the component of V in the direction* $\vec{u}$ *by computing the "dot product" or "inner product"*

$$\vec{u} \cdot \vec{V}$$

Another tag says: *if you have any non-zero vector* $\vec{W}$*, you can get a* unit *vector* $\vec{u}$ *by dividing* $\vec{W}$ *by its length.*

Yet another tag says: *if you have the equation of a plane P in the form*

$$a\,x + by + cz = d,$$

where $a^2 + b^2 + c^2 > 0$, *you can immediately write down a vector* $\vec{N}$ *that is normal to plane P.*

Carrying out such advance planning of how to attack a problem depends upon:

i) knowing the necessary techniques

ii) possessing appropriate descriptors ("tags" or "labels") for each technique, that specify what the technique can accomplish

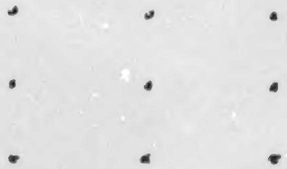
iii) (probably) possessing a definite collection of recognizable sub-goal *candidates* (that is to say, a "menu" of *possible* sub-goals from which appropriate sub-goals for a given problem can be selected) or a way to synthesize them

iv) mechanisms for identifying appropriate sub-goals and retrieving the appropriate "tags" or "labels"

v) given a "tag" or "label", a mechanism for retrieving its associated tool

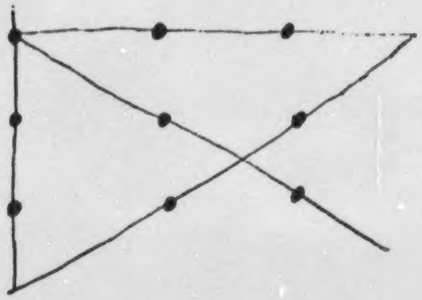
vi) mechanisms for assigning correct inputs for each "tool" or sub-procedure.

It is sometimes assumed that a problem solver has, laid out in his or her mind, a complete "tree " of possibilities - of *all* the possibilities, that is. There seems to be no observational data in support of such an assumption. On the contrary, people commonly "see" (or "bring to mind") very *few* possibilities, and may, indeed, omit the most promising ones. (This happens, for example, in the puzzle: draw a sequence of connected straight line segments without lifting pen from paper, so that each dot lies on at least one of the segments; do this with the smallest possible number of segments:



;

The smallest number turns out to be 4, as in this solution, which most people do not consider.)



)

The search problem may seem simple, if one unconsciously assumes the student knows just where to look. But — for the "distance to the plane" problem — consider all of the things the student *could* do, such as

i) form the vector $\overrightarrow{AB} = \overrightarrow{(-a, b, o)}$ from the point A (a, o, o) to the point B (o, b, o);

ii) form seven times the vector $\overrightarrow{AB}$:

$$7\,\overrightarrow{AB} = (-7a, 7b, o) \qquad ;$$

iii) form the cross product

$$\overrightarrow{AB} \times \overrightarrow{BC} \qquad ;$$

iv) find the unit vector

$$\frac{\overrightarrow{AB}}{\|\overrightarrow{AB}\|} \qquad ;$$

v) form the triple scalar product

$$(\vec{A} \times \vec{B}) \cdot \vec{C} \qquad ;$$

vi) find the distance from point A to point B;

and so on.

The number of possibilities is clearly infinite; but even when the number of possibilities is finite but large, people do not typically recognize most of them, and may omit some of the most important. What guides good problem solvers to "grow" the search tree in the most useful directions?

A further example of planning in advance of calculation: Suzuki (1979), when a student in eleventh grade calculus, decided to solve this problem by as many different methods as she could devise:

The curve C is defined by the equation

$$5x^2 - 6xy + 5y^2 = 4.$$

Find those points on C which are nearest to the origin.

Among the methods Suzuki concocted were:

i) recognize this as the equation of an ellipse whose axes are rotated in relation to the coordinate axes; therefore rotate the coordinate axes to coincide with the axes of the ellipse, and read off the semi-major and semi-minor axes by inspection;

ii) let $\ell^2 = x^2 + y^2$ be the square of the distance from the origin to the point $P(x, y)$. Minimize ℓ^2, subject to the constraint that the point P must lie on the curve C, by getting a pair of simultaneous equations in the differentials dx and dy, and use the Cramer's rule requirement that the determinant of the coefficients must be zero;

iii) the vector $T = (1, y^1)$ is tangent to the curve C; it must be perpendicular to the vector $\vec{R} = \overrightarrow{(x, y)}$ from the origin to the point of tangency; therefore set the dot product equal to zero: $\vec{R} \cdot \vec{T} = 0$;

iv) introduce a mythical "temperature" T, defined as $T = 5x^2 - 6xy + 5y^2 - 4$. The gradient $\vec{\nabla} T$ points in the direction in which T increases most rapidly, and is normal to a curve of constant temperature. At the nearest point, the vector $-\vec{\nabla} T$ must point directly toward the origin, hence be parallel to (and therefore a scalar multiple of) the vector $\vec{R} = \overrightarrow{(x, y)}$;

v) convert to polar coordinates, and minimize r.

The specific planning by which Suzuki created these strategies was not reported; nonetheless, it seems clear that she possesses a very powerful capability for planning novel ways of solving problems. (More recently, she has made excellent showings in various problem-solving contests.)

Yet one more method for solving this same problem has been employed by Kumar (1980), while a student in twelfth grade calculus: examination shows that C is a smooth curve contained in an annular ring centered at the origin. Therefore, a very small circle

$$x^2 + y^2 = r^2 ,$$

where r is small, will not intersect C. But as r becomes larger, the circle will intersect C. Finally, for still larger r, the circle will be outside the annular ring, and will not intersect C. Hence, find the smallest positive value of r such that the system of simultaneous equations

$$\begin{cases} 5x^2 - 6xy + 5y^2 = 4 \\ x^2 + y^2 = r^2 \end{cases}$$

has a solution in real values of x and y. One can set out to solve this system; if the result is a quadratic equation, the criterion $b^2 - 4ac = 0$ will identify the desired value of r.

So much for *strategic* planning; when this strategy is implemented, the result is not a quadratic equation, but rather a fourth-degree equation. With a little ingenuity, a tactical step can be inserted that transforms the fourth-degree equation to a quadratic, and the problem is then easily solved.

Finally, one can look at planning as it is carried out by more experienced problem solvers. R, an experienced calculus teacher, was planning a lesson on writing the equations of tangents and normals to various curves. He attempted to sketch out his *a priori* plans, to the extent that he was aware of them : (Note 6)

"I thought of two things: (i) distinguish the point (x,y) [the "general" point that moves along a curve, or tangent line] from the point (x_1,y_1) [the fixed (or "constant") point of tangency]; and (ii) use two geometric criteria to get two equations - the tangent and curve must *intersect* at (x_1,y_1), and they must *have the same slope* at (x_1,y_1).

"Later on, as I began to think more seriously about the problem, I wanted to write equations for the pencil of lines through the point $P\ (\frac{3}{2}, 0)$, and to do this I selected the form

$$\frac{y-0}{x-\frac{3}{2}} = \frac{y_1-0}{x_1-\frac{3}{2}} .$$

"Still later, when I saw C, [another mathematics teacher in the same school] use the point-slope form of the equation, that struck me as more natural, and so I switched to using that form."

Notice especially the use of an appropriate and rather well-developed *meta-language*, with terms such as "get two equations", "the 'general' point", "the fixed point of tangency", "the pencil of lines", "the point-slope form of the equation", "struck me as more natural", and so on. (To see a remarkably sophisticated meta-language used by a seventeen-year-old high school student, cf. Parker, 1980.)

In addition to mathematical tasks, the study also used certain puzzles, particularly the "cannibals and missionaries" puzzle, and a word puzzle (start with "SOUP", and change only one letter at a time, every line being a legitimate word, to end up with "NUTS").

The best of the eleventh-graders were exceptionally skillful in

(i) setting goals and sub-goals

(ii) using a powerful meta-language to describe and analyze what they were doing

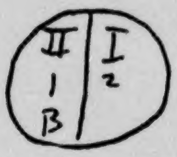
(iii) making very quick revisions in their strategy when they got a glimpse of a new possibility, or when they saw a dead-end looming ahead.

Here are excerpts from a transcript of an interview with Witold, a highly superior problem solver (who was 15 years old). Witold is so good at mathematics that his performance on mathematical tasks is often hard to learn from -- it all seems so effortless, and happens so quickly. His problem-solving technique is revealed more clearly in the puzzle protocols.

In the "Cannibals and Missionaries" puzzle, you have three missionaries and three cannibals on the left bank of a river. They want to cross, and have a boat that can be rowed by one person, but can hold at most two people. If at any time, on either bank, the cannibals outnumber the missionaries, a disaster occurs. Your task is to work out a sequence of non-catastrophic crossings that will get all six people across the river.

Here is [illegible]old, working on this puzzle, which is entirely new to him:

1. [First of all, he spontaneously and immediately inverts a notation to describe the "state" of the system:



would mean:

on the left bank	on the right bank
2 missionaries	1 missionary
1 cannibal	2 cannibals
the boat	

2. [Second, he chose to consider [illegible]ore-or-less) every possible pattern of crossings, using a tree diagram to help him keep track.]

3. Here is how he proceeded:

Student: You could do that just by drawing a tree like...OK. You've got...use Roman numerals to represent missionaries and numbers to represent cannibals, and your starting position is III, 3 and a B for the boat, and then from there you've got, what did I say? Roman numerals are cannibals?

Interviewer: ...Missionaries.

Student: OK, fine. Roman numerals are missionaries so then you have 2 options here. That doesn't get anybody eaten and that's · you can either go to having one, three, and two, you see my notation system here, the line is the river.

Interviewer: OK

Student: Or you could go to...that's what I did there, two, and then symmetrically you could do the same thing from the other end, so if you ever hit a symmetrical position, you've got it made and, OK, but this one you see is a dead end because somebody has to come back. It's just like what you could have gotten from over here, so if the cannibal comes back, then these people get eaten, so again, all this can lead to is stuff you can get from over here, anyway. So you can forget about that one. This one could continue. You could either bring everybody back but that's obviously silly. You're back to where you started, or, and these trees on a simple problem like this won't spread out because things keep getting eliminated. So then the only thing you could go there from here, I guess, is having two in a boat and a cannibal over there and then from here we only have one choice. You can send over one cannibal but again that's putting you back to there. You can send over one missionary which puts you back to there, so you've gotta send over two. And sending over both missionaries loses the third missionary, too. You send over both cannibals and then you've got 3--and that, now at the point if you can then get--you've got to hit 2 symmetrical positions, for example, if you hit this and this with the boat on the other side, which is impossible. Am I making a fool of myself here? I guess from here the only sensible thing,you could just keep brute forcing on through. Like from here the only sensible thing sending both cannibals back across would put you back up there so from here the only thing that works is, two, the boat, no, no boat, two, one, three, boat. Now at this point you can send the canni--if you send anything but 2 missionaries back, you'll either get something you've been at before or get somebody eaten, so going back up to there what you have is you've got one of each over here and then over here you've got two, and two, and the boat, 'cause anything else would be a repetition and so on. From there it's almost done. You want me to finish it?

Interviewer: Well, it's clear what you do. Now one of these is a Roman numeral.

Student: Right.

Interviewer: And so at this point one of each has to go back.

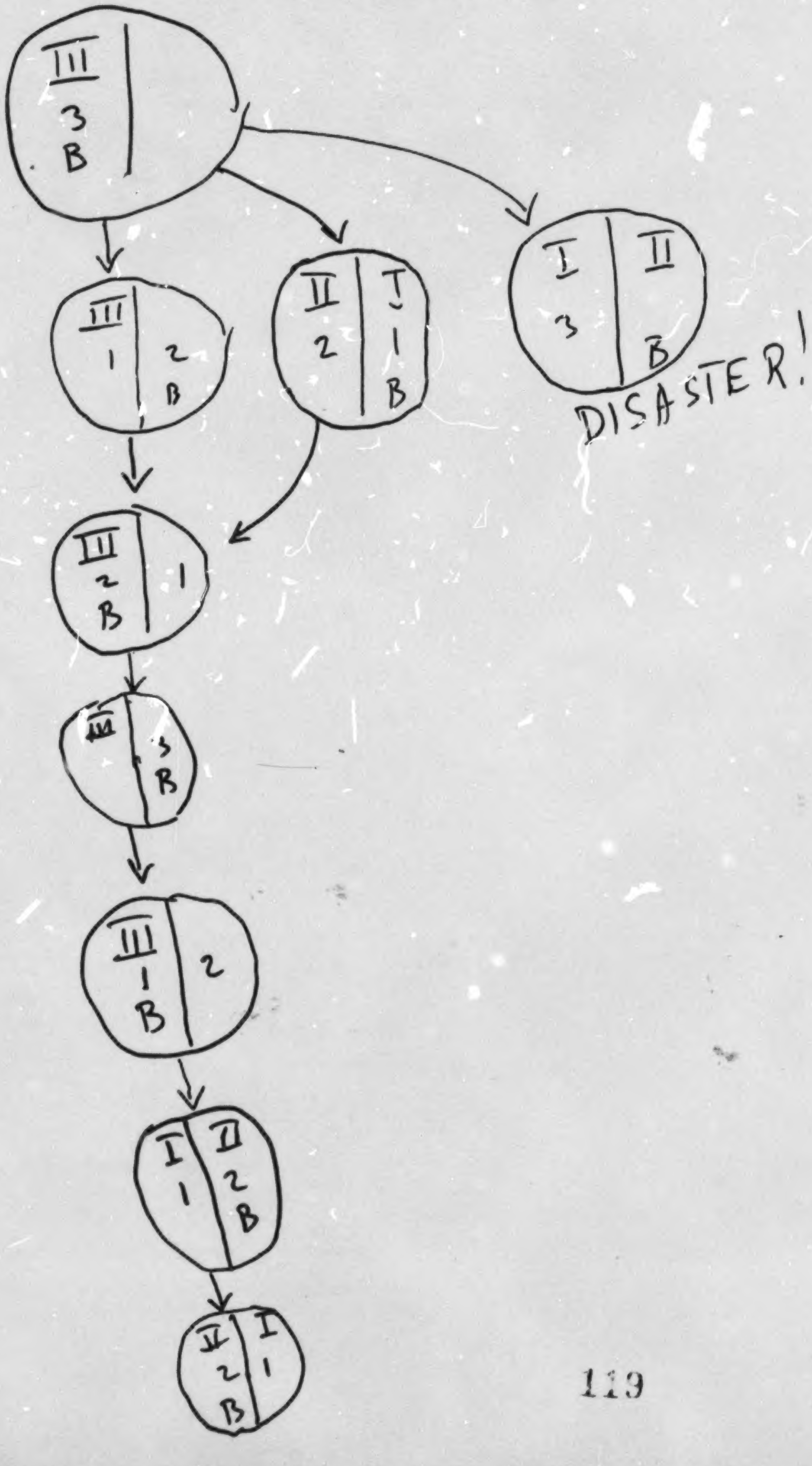
Student: No, wait, sending back one of each. Right, sure it will. If you send back one of each, then you can follow these steps in reverse order.

Interviewer: Yeah, right. If you send back one of each, then two missionaries can now come back next time.

Student: Well, if you send back one of each, you just follow these in reverse order... Which is just like this except the sides of the river have been reversed and what you want is this with the sides of the river reversed so if you reverse all crossings...

Interviewer: Aha, nifty! Oh wow! That's phenomenal! Oh wow!

4. Here is what he wrote as he was saying that:


III 3 B
III 1 2 B
II 2 I 1 B
I 3 II B
DISASTER!
III 2 B 1
III 3 B
III 1 B 2
I 1 II 2 B
II 2 B I 1

5. [Notice W's stunning use of meta-language, to describe and analyze what he's doing -- especially his criterion that if a "symmetry" can be found, the problem is solved! And his stating a rule to proceed after a symmetry has been found: reverse both the time sequence, and the river!]

XVI. *Summary.*

For most adults in past years, "mathematics" has probably meant memorizing certain specific techniques, and thereafter recalling them and using them as necessary. This is probably still true for many people even today. For a minority of people - including mathematicians, engineers, scientists, computer specialists, and a growing number of people in the health care field, in psychology, in education, in art, and elsewhere - the use of mathematics has nowadays a different quality. For these people, mathematics is an expandable tool for solving novel problems, for which no previously-learned algorithm will be entirely sufficient. An engineer designing a more fuel-efficient automobile engine, or a composer using a computer to generate a new piece of music, is not merely plodding along by stepping in someone else's footprints. He or she is exploring new territory - and this often means exploring new *mathematical* territory, as well. Hence thoughtfulness, the re-thinking of old assumptions, intelligent guessing, insight, and shrewd planning are all integral parts of the task.

Unfortunately, this "creative" aspect of mathematics has typically been ignored in most school curricula, and the whole idea is foreign to most of the general public. Something very important thus gets lost. In order to correct this situation, we must devote more effort in school to teaching the creative aspects of mathematics, and in our research and development work as well we need a greater emphasis on creativity, understanding, and problem solving.

In recent years, an alternative research paradigm has appeared in the world of mathematics education. This alternative paradigm shows considerable promise for increasing the emphasis on creativity, vision, understanding, insight, and the like, while at the same time paying proper attention to rote drill, routine practice, and "meaningless" algorithmic performance. This alternative paradigm gets data especially from task-based interviews, but also uses other data-collection methods, including the analysis of error patterns, and even the precise measurement of response times. In typical cases, the analysis of this data is based upon information-processing conceptualizations, often drawn from the fields of cognitive science and artificial intelligence.

This approach is beginning to demonstrate its ability to create a serious theory for the analysis of those processes of thinking that are required in dealing with tasks of a mathematical nature.

It has been a most gratifying aspect of this research project that, simultaneously with the Project work in Urbana, Illinois, an unprecedented outpouring of relevant work has been underway, independently, by many other investigators, especially the Clement-Lochhead group at the University of Massachusetts, Matz at M.I.T., John Seely Brown and his colleagues at B.B.N. and at Xerox, Karplus, Stage, and their colleagues at U.C. Berkeley, the Minsky-Papert group at M.I.T., Roger Schank and his colleagues at Yale, Steffe and colleagues at Athens, Georgia, a large group of investigators in Pittsburgh, a group in San Diego, Kristina Hooper's group at the University of California at Santa Cruz, Robert Lawler and Andrea DiSessa at M.I.T., and by many others elsewhere.

Mathematics education has not often seen such a period of rapid development, in similar directions, by so large a number of independent investigators. What is developing is a shared view of mathematics learning that is quite different from the typical view of, say, the 1950's. This gives us an entirely different way of thinking about the processes of mathematical thought. In SECTION FOUR we use this conceptualization in order to restate the results of this study of "understanding" in mathematics.

SECTION FOUR

In this section we reinterpret our observational studies in terms of the information-processing concepts of the preceding section.

XVII. Information-Processing Aspects of "Understanding," or of a Failure to Understand.

A. Matching Input Data to a Structure Retrieved from Memory.

In any case, "understanding" requires the use of a knowledge representation structure (krs) within the student's mind. This krs can be retrieved from memory, it can be constructed, or it can be created by a combination of retrieval and new synthesis.

In the case where retrieval is involved, one would ordinarily be said to "understand" if all of the following are true:

1. A knowledge representation structure exists in the student's memory which can adequately represent the present problem situation;

2. An appropriate krs is selected and retrieved from memory;

3. Specific input data from the present problem is mapped correctly into the variable slots in the retrieved krs;

4. When the task goes beyond the specific krs, the necessary additional krs's are activated (this is, in effect, a recursive return to the first information-processing task).

Thus, the teachers in our study might reasonably be said to understand a task such as:

$$\text{Find } \frac{dy}{dx} \text{ if } y = 3x^2 + 5x - 7.$$

Each of the four conditions is easily seen to be satisfied.

For the students this was not necessarily the case. One easily finds instances where one or more of the four conditions were not satisfied -- as, for instance, where associations to the fundamental definition

$$f'(x) = \lim_{h \to 0} \frac{f(x + h) - f(x)}{h} \qquad (1)$$

were not activated, or where the further associations connecting (1) to the slope of a secant were not activated. [This last matter is easily tested in an interview format by asking the student to give a graphical interpretation of (1).]

B. Construction of a Representation.

A knowledgeable student possesses many powerful devices for constructing representations. One of the most important is the use of Cartesian coordinates, as in explaining the distinction between

$$\lim_{h \to 0} \frac{1}{x} \log_a \left(1 + \frac{h}{x}\right)^{\frac{x}{h}}$$

and

$$\frac{1}{x} \log_a e ,$$

where in the one case a limit is taken along the x-axis, and in the other case a limit is taken along the y-axis. This example uses also another powerful representation device: the sequential order in which operations are carried out.

A student who cannot retrieve a suitable representation, and who also cannot construct one, would not be said to "understand."

C. Representations constructed by a combination of retrieval and ad hoc construction. This process typically requires a breaking up of the present problem into several parts or segments, with krs's being retrieved for individual parts, and "welded together" to represent the entire problem. (Note 7) Being able to identify the key "critical points" is a measure of the degree of "understanding."

D. Associations.

Clearly, any serious depth of understanding will usually require a rich network of appropriate associations (i.e., connections to other krs's). At least four mechanisms have been postulated to account for associations. (Cf. Davis 1982-B.) They can be summarized briefly as:

1. "Sending" or "pointers": the operative krs may contain pointers that direct processing to the appropriate further krs;

2. "Volunteering" or "pattern recognition": the products of intermediate information-processing may be treated much the way that initial input data is, and be scrutinized for certain patterns whose appearance will trigger the activation of other krs's.

3. Both of the preceding mechanisms may be combined, as in some of the work of Gerald deJong;

4. The "library-card" model. Consider the act of going to a library, examining a particular book, looking to see who has previously checked it out, and checking it out yourself if certain other people have previously used it. This can serve as a metaphor for a mechanism proposed by Schank and Kolodner: since it is

widely assumed that what is stored in memory includes a record of which procedures and representations have been used, one can use this record to locate likely associations: if some other process used this same item, examine that process as possibly relevant to our present task!

E. Recognizing the Over-All General Nature of the Task.

We considered earlier Anne's study of Laplace transforms. Anne did not recognize the "general nature of the task," which could be described as follows:

There are two sets of mathematical entities:

$$S_1 \qquad\qquad S_2$$

There is a correspondence between elements of S_1 and elements of S_2:

$$\begin{tikzcd} S_1 \arrow[r, leftrightarrow, "\overrightarrow{T}", "\overleftarrow{T^{-1}}"'] & S_2 \end{tikzcd}$$

There are some calculations that are supposed to be carried out on elements of S_1...

but, instead, we map S_1 into S_2,
carry out some (different)
calculations on S_2, and
then map back into S_1.

Note that this description applies equally well both to logarithms and to Laplace transforms.

If one is unaware of the general nature of the tasks, and tries to proceed by following various formulas, one is proceeding by rote without "understanding" (and one is very likely to fail).

This is a very interesting phenomenon from an information-processing perspective. One of the fundamental properties of information-processing conceptualizations is that information, and information-processing, can exist on a hierarchy of levels: what on one level is a process may be looked at, or operated upon, at a meta-level, where it will be treated as a piece of data. Hence, just as one can see the pattern that unites

$$(x - 2)(x - 3) = 0$$

with

$$(x - 5)(x - 1) = 0$$

and with

$$(s + 7)(s - 3) = 0$$

so also one can see that using

$$\log A + \log B = \log AB$$

follows the same pattern as using Laplace transforms to solve an initial value problem for a linear differential equation with constant coefficients.

The "general nature of the task," then, becomes merely a matter of pattern-recognition, except that the patterns are now to be seen from a higher "meta" level. Of course, the experience that a student needs in order to be able to operate on these higher "meta" levels is quite different from that which is required for operating on the lower "factual" levels.

F. Setting Sub-goals; "Following the Story Line."

We have seen that every step in working on a problem is aimed at achieving some goal or sub-goal. Thus, we derive the quadratic formula by taking steps such as:

$$ax^2 + bx + c = 0, \quad a \neq 0$$

$$x^2 + \frac{b}{a}x + \frac{c}{a} = 0$$

[The goal of this step was to make the coefficient of x^2 one, in order to simplify the process of completing the square.]

$$x^2 + \frac{b}{a}x = -\frac{c}{a}$$

[The goal of this step was to make it easier to complete the square.]

$$x^2 + \frac{b}{a}x + \frac{b^2}{4a^2} = -\frac{c}{a} + \frac{b^2}{4a^2}$$

[Completing the square; the goal was to get to an equation form that will factor easily.]

$$\left(x - \frac{b}{2a}\right)^2 - \left(\frac{b^2}{4a^2} - \frac{c}{a}\right) = 0$$

[We've finally achieved a form that factors!]

$$\left[\left(x - \frac{b}{2a}\right) + \sqrt{\frac{b^2 - 4ac}{4a^2}}\right]\left[\left(x - \frac{b}{2a}\right) - \sqrt{\frac{b^2 - 4ac}{4a^2}}\right] = 0$$

[The purpose of this last step was to achieve a form where we could use the Zero Product Principle.]

A major ingredient in "understanding" is being able to "follow the story line" -- to recognize sub-goals, or even to set sub-goals for oneself.

Notice the parallel between "recognizing the general nature of the task" and "setting sub-goals" -- both, viewed from a higher "meta" level, are essentially pattern-recognition tasks.

G. There is the somewhat different category, discussed earlier, of "knowing what you yourself are doing." This, too, can be restated in information processing terms, but we do not take the space here to do so.

H. Intuition: The Primitive Foundation Schemas.

What is "mathematical intuition"? The emerging view appears to be that "intuition" is related to the process by which early knowledge representation structures (or "schemas") become elaborated with experience. In particular, this elaboration process often starts with very simple schema, dealing with the kinds of experiences that very young children have, experiences such as turning one's body, moving a step forward, putting one object near another, drawing a curved "line," etc. An early schema, for example, that relates to giving each child one cooky, or putting hats on dolls so that each doll gets one hat can, as one advances in mathematics, become the basis for great elaboration, getting later to matters such as

$$F(s) = \int_0^\infty e^{-st} f(t)\,dt\,,$$

which establishes a mapping between functions, f (t), and their Laplace transforms, F(s), as in:

$$f\,(t) \longleftrightarrow F\,(s)$$

$$e^{-at} \longleftrightarrow \frac{1}{s+a}$$

$$\sin t \longleftrightarrow \frac{1}{s^2+1}$$

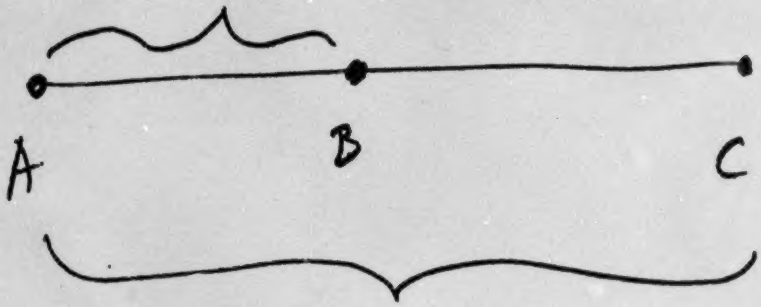
$$\cos t \longleftrightarrow \frac{s}{s^2+1} \quad ,$$

a correspondence having much in common with putting hats on dolls so that each doll gets exactly one hat.

If secure schema exist, and have been carefully elaborated, then we have an "intuitive" grasp of a mathematical situation. Otherwise we do not.

I. Schemas Without Antecedents.

As discussed earlier, the importance of extended elaboration of schemas, starting with *very simple* initial schemas, is revealed most strikingly in mathematical situations for which *there are no* simple antecedents. Consider the theorem that the number of points on segment [A, B] is *exactly* equal to the number of points on segment [A, C]:



When this theorem was proved to him, Robbie (in grade 12) concluded that somehow the points on segment [A,B] were being "squashed together tighter" -- although, of course, there are obvious problems with any such interpretation. Points simply do *not* behave like ball bearings, or buck shot, or moth balls, or rubber pellets, or grains of sand, or any other *physical* objects.

Those who study mathematics must, as many of them say, "educate their intuition." With the advent of space travel, we all needed to "educate our intuitions" about the effects of *air*. A moon vehicle, returning to earth at the wrong angle, would "skip" off the earth's atmosphere and travel on into space, and as a flat stone can "skip" across a pond if it is thrown in a special way. But a space capsule -- skipping off of the thin, insubstantial atmosphere? It seems impossible. That means we must educate our intuitions, and realize if you hit the earth's atmosphere at a high enough speed, it can be like hitting water!

The following interesting example was provided by Oliver Selfridge and Edwina Rissland: we think of a sphere in three-dimensional space as round and compact. A three-dimensional *cube* is not quite so round and compact as a sphere, but it comes close -- it is *reasonably* round and compact. Consider, now, a cube in a space of 100 dimensions. Say the cube has one corner at the origin, and is one unit along each edge. The cube is then

$$C = \left\{ (x_1, x_2, \ldots, x_{100}) \;\middle|\; 0 \leq x_i \leq 1,\ i = 1, 2, \ldots, 100 \right\}$$

Is this 100-D cube also "reasonably round and compact"? By no means! It must be thought of as sharply pointed, with spear-like projections at every corner. Why? Consider the distance from the center of any face to the center of any other face. This distance cannot be more than 1, because one can go to the midpoint of the edge where the faces meet by a journey of $\frac{1}{2}$, and traveling via this point one can go from the center of one face to the center of the other by a journey whose length is $\frac{1}{2} + \frac{1}{2}$. (There may be shorter routes!) But now go from one vertex to a "diagonally" opposite vertex. In 2-D, the distance is $\sqrt{2}$. In 3-D, the distance (as one easily figures out) is $\sqrt{3}$. In dimension N, the distance is $\sqrt{N}$. In 100 dimensional space, the distance is $\sqrt{100} = 10$. So the faces are "all near one another," yet -- somehow -- the vertices "stick way out!" "Understanding" this requires some serious work on "educating your intuition."

But, in terms of early "basic" well-established knowledge representation structures, we have powerful capabilities for three dimensional space -- watch professional basketball players! -- but no experience at all in E_n for $n > 3$. Our schemas for E_{100} must be developed by carefully deliberate elaboration, from the quite different schemas for E_3.

J. Misinformation.

One does not necessarily approach a topic with "no idea at all." On the contrary, one may come to a new topic laden with a heavy burden of incorrect ideas.

Within our studies, this phenomenon has been especially conspicuous in relation to the concept of limit of a sequence. Among the persistent wrong ideas that students have repeatedly used are these:

1. The sequence 1, 1, 1, ... does not converge. In a convergent sequence the terms are getting nearer to the limit -- but these terms aren't "getting nearer to" anything!

2. The limit of the sequence .9, .99, .999, .9999, ...

is a number "just a little bit smaller than one." [This is an interesting error -- it is of a very common type, errors due to a failure to distinguish two different things.] These students are correct in thinking that the terms of the sequence "will always be a little bit less than one." The students err, however, in failing to distinguish between

the terms of the sequence

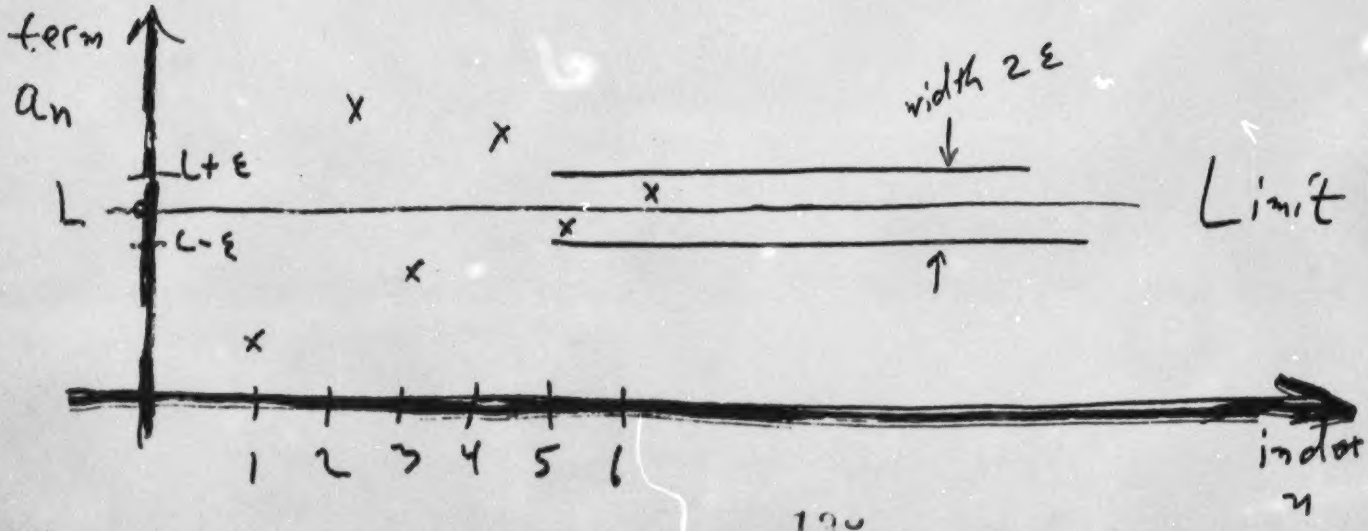
and

the limit of the sequence.

Minsky's (and Scott Fahlman's) notion that

> "a general frame will [often] use one of the specific cases below it as its exemplar; 'mammal' might simply use 'dog' or 'cow' as its exemplar, rather than trying to come up with some schematic model of an ideal nonspecific mammal " (Minsky, 1975, p. 266)

can be applied to postulate a description for the representation structures involved. What should have been constructed as an appropriate representation structure would provide for something like (or at least equivalent to) the familiar graphical sketch,



which shows the (admittedly complex) relationship between the terms a_n and the limit L.

The students who make the present error are, presumably, not making use of such a representation, but are instead using some term, a_n, for reasonably large n, as an "exemplar" representative of the limit. "Cow" may (depending upon context) suffice as the representing exemplar for "mammal," -- and, indeed, a_n will even suffice, in some computational contexts, as a representing exemplar for L -- but in theoretical discussions the distinction can become crucial!

3. Some students imagine, incorrectly, that every convergent sequence is monotonic, and are thus led to make many inappropriate definitions, and to draw many false conclusions.

4. The notion that, for $n < N$, the initial terms a_n will be disregarded is difficult for some students to accept. These students appear not to distinguish between a series such as

$$1 + \frac{1}{4} + \frac{1}{9} + \frac{1}{16} + \ldots + \frac{1}{n^2} + \ldots$$

and a sequence such as

$$1, \frac{1}{4}, \frac{1}{9}, \ldots, \frac{1}{n^2}, \ldots$$

(although their text and their teachers have attempted to avoid such confusion). These students need to switch basic metaphors; instead of thinking in terms of something like a savings account in a bank, where previous deposits accumulate and continue to make their presence felt in the current balance, they need to think of a metaphor such as a calculation -- if one calculation is rejected as too imprecise, a new calculation, somewhat more precise, is carried out. Earlier, less precise, calculations are, of course, discarded.

K. Critics

A "critic" is a procedure that checks up on a frame retrieval, or on instantiation of variables, or on some other process, and pronounces it "acceptable" or "unacceptable." Thus, a size critic ought to have told Marcia that

$$\begin{array}{l} 5 \\ \cancel{6} \\ \cancel{7},\ 0\,{}^{1}0\,{}^{1}2 \\ \quad\ -\ 2\ 5 \\ \hline 5,\ 0\ 8\ 7 \end{array}$$

could not be correct, since the answer had to be close to seven thousand. People who "understand" a subject well seem to have many critics available, so that (for example) they do not say

$$dy = \sec^2 x \quad , \qquad (1)$$

nor

$$y' = (\sec^2 x)\, dx, \qquad (2)$$

because neither (1) nor (2) is of the correct form with regard to differentials. Beginning students, however, make such errors frequently.

L. Knowledge of Examples

As work by Rissland and others has made clear, "understanding" is often related to the matter of knowing a large number of examples. One might, for example, come to the erroneous conclusion that a set which is not closed is open, unless one knew examples such as

$$\left\{ x \mid 0 \le x < 1 \right\} \ .$$

M. Levels

We have referred, above, to various "meta" levels. One might speak of a "performance" level, then a "meta" level that observes, analyzes, and guides the performance-level procedures, then a "meta meta" level that modifies the meta level, etc.

Suppose the problem were to divide two fractions:

$$\frac{3}{5} \div \frac{2}{7} =$$

A "performance-level" procedure might invert the first fraction, then multiply:

$$\frac{5}{3} \times \frac{2}{7} = \frac{10}{21} \quad .$$

If no meta-level procedures were in operation, this might stand as the result. If, on the contrary, appropriate meta-level procedures were activated (perhaps by a warning signal that dividing fractions could be tricky -- better check!), then for example, a certain powerful heuristic procedure might be invoked:

Try a simple case where you know the answer!

[This, for example, is how many people test a calculator.]

Now,

$$\frac{1}{2} \div \frac{1}{4}$$

has, because of its meaning, the answer "2":

$$\frac{1}{2} \div \frac{1}{4} = 2.$$

But the procedure just used would have given an incorrect result:

$$\frac{2}{1} \times \frac{1}{4} = \frac{1}{2}$$

(Even more ingeniously,

$$4 \div 2 = \frac{4}{1} \div \frac{2}{1}$$

which, using the procedure above, would become

$$\frac{1}{4} \times \frac{2}{1} = \frac{2}{4} = \frac{1}{2},$$

whereas $4 \div 2 = 2$.)

This "critic" procedure thus reports an error. The "invert" procedure is not working! An appropriate meta-procedure needs to intervene, and to modify the "invert" procedure. For example, this meta procedure might operate by changing the *choice* of which fraction to invert, leading to

$$\frac{a}{b} \div \frac{c}{d} = \frac{a}{b} \times \frac{d}{c},$$

which would lead to

$$\frac{1}{2} \div \frac{1}{4} = \frac{1}{2} \times \frac{4}{1} = \frac{4}{2} = 2$$

a correct result! Also,

$$4 \div 2 = \frac{4}{1} \div \frac{2}{1} = \frac{4}{1} \times \frac{1}{2} = \frac{4}{2} = 2$$

also correct! Using this *modified* performance procedure, one gets

$$\frac{3}{5} \div \frac{2}{7} = \frac{3}{5} \times \frac{7}{2} = \frac{21}{10}$$

(which, of course, is actually correct).

Consider the following problem, on which we have a very large amount of data: (Note 8) students in *arithmetic*, who have not studied algebra (and have NOT learned standard techniques for solving quadratic equations), are asked to solve these equations

$$(\square \times \square) - (5 \times \square) + 6 = 0$$

$$(\square \times \square) - (14 \times \square) + 33 = 0$$

$$(\square \times \square) - (7 \times \square) + 10 = 0$$

$$(\square \times \square) - (12 \times \square) + 35 = 0 \quad ,$$

and others, where the solutions are, in every case, unequal primes.

What procedures can the students use? At first, only trial-and-observation. By trial-and-observation, students easily find that the solutions for the first equation are 2 and 3, for the second they are 3 and 11, for the third they are 2 and 5, and for the fourth they are 5 and 7. Most fifth graders soon realize that there is a "shortcut" -- try factors of the last number! I.e.,

$$6 = 2 \times 3$$
$$33 = 3 \times 11$$
$$10 = 2 \times 5$$
$$35 = 5 \times 7 \quad .$$

The discovery of this shortcut is not the work of an ordinary "performance-level" procedure. If one postulates a hierarchy of "meta" levels of procedures, presumably one would assign the shortcut-detecting procedure to a higher meta level than that of the regular "operational" procedures for carrying out assigned arithmetic operations.

After students have discovered the shortcut, they are presented with the equation

$$(\square \times \square) - (13 \times \square) + 40 = 0 \ .$$

The majority response is to claim (without substituting for confirmation) that this equation has six solutions, namely 10, 4, 8, 5, 20, and 2. When students are urged to check, they find that in fact only 8 and 5 are solutions! Those students who spontaneously discover that "there are two rules" have presumably again used a meta-level procedure to discover that, for the equation

$$(\square \times \square) - (a \times \square) + b = 0 \,,$$

the solutions r and s must satisfy the two rules

$$r + s = a$$

$$r \times s = b$$

We do not propose any specific mechanism for assigning procedures to specific meta-levels. Among procedures for which one might seek such an assignment, consider:

a) A procedure that employs the heuristic: "If you find one 'secret rule,' don't stop looking! There may be other 'secret rules' as well!"

b) A considerably more general version of the preceding procedure, which employs the heuristic: "Whatever you are doing, if an amount x of it seems to succeed, try a larger amount (or more instances) of it. Maybe more will be better!"

c) A far more specific procedure that employs a heuristic: "If the product of r and s turns out to be important, check for patterns that involve the sum $r + s$, the differences $r - s$ and $s - r$, and the quotients $r \div s$ and $s \div r$."

d) A procedure concerned with establishing a connection between the observed patterns and the axiomatic logical structure of algebra.

e) A procedure concerned with the question of why the first four equations seemed to have two solutions (an appearance corresponding to fact), whereas the fifth equation presented at first the deceptive appearance of having six different roots. (The answer, of course, is that the roots of the first four equations were unequal primes, so the factorization of the constant term seemed unique. By contrast, $40 = 2 \times 2 \times 2 \times 5$.)

Even without any precise specification of "meta" levels, it is apparent that different students "understand" this quadratic equation problem in quite different ways.

N. Different Kinds of Knowledge.

One may "know" in many different senses. For mathematical purposes, it is important to distinguish at least these (Davis, 1982-B):

i) "Knowledge" in the sense of possessing the ability to retrieve and execute an appropriate algorithm.
ii) "Knowledge" in the sense of seeing relations to basic or "primitive" schemas.
iii) "Knowledge" in the sense of the ability to retrieve a frame which, in turn, is part of a well-elaborated frame system (cf., e.g., Minsky, 1975, pp. 216-219).

SECTION FIVE

IMPLICATIONS FOR INSTRUCTION

XVIII. Three Instructional Corollaries.

A. The "Paradigm" Teaching Strategy.

One of the most important (and least recognized) aspects of the "new mathematics" was a method for introducing new problem situations, or new concepts, in strikingly concrete forms. The concept of surface area might be introduced in terms of the task of coloring a block by stamping with a square stamp 1 cm.

by 1 cm. -- how many times must you stamp? Negative integers might be introduced by putting pebbles into, and taking pebbles out of, a bag that was already partly filled with pebbles. The concept of isomorphism might be introduced by successive modifications of the game of Tic Tac Toe. The inner product of vectors might be introduced by computing the amount of money spent in a candy store. (Davis, 1980-A)

At the time that this style of concrete presentation of new ideas was developed -- primarily from 1956 to 1970 -- no theoretical justification was offered. The emphasis which recent cognitive science studies have given to the process of building elaborated frames on a foundation of basic "primitive" frames provides a theoretical conceptualization within which the use of concrete introductions of new ideas becomes understandable.

B. "Discovery" Teaching.

If the technique of the concrete introduction of new ideas was a little-noticed aspect of the "new mathematics" curricula, the same can hardly be said of "discovery teaching." "Learning by discovery" became a commonplace slogan for most major publishers, was the subject of several scholarly conferences (cf. Shulman and Keislar, 1966), and was even presented on network television during prime time. Yet the discussion of "discovery learning" in the 1960's was generally handicapped by the incorrect assumption that the goals of expository teaching and of discovery teaching were the same. They were not, but the distinction can be stated more clearly in terms of modern cognitive science conceptualizations. For a large proportion of expository teaching, the goal is the creation of specific "performance level" procedures. Discovery teaching aimed also at the development of meta-level procedures -- indeed, emphasized meta-level procedures, on the grounds that specific algorithms could be learned when they were needed (as, indeed, they often must be).

If the mechanisms postulated by Schank and Kolodner, among others, are in fact prominent -- especially storing in memory a record of procedures that have been used -- then expository teaching and discovery teaching cannot be teaching the same thing, because they are not laying down the same memory traces. And, if meta-processes are important to creative and flexible performance, there is the danger of a very considerable loss when highly-explicit (often rote) expository teaching is employed.

C. Task-Based Interviews: Probing the Student's Understanding.

The "new mathematics" movement of the 1960's made little use of task-based interviews, despite the fact that Piaget had been using this method brilliantly for decades. Interviews to probe more deeply into the way that a student understands a topic (or thinks about the topic) gained acceptance within the United States primarily as Erlwanger, Clement, Lochhead, Karplus, and others gave convincing demonstrations of what this method could accomplish. The value of such interview techniques is, nowadays, hardly open to question. Without such probes, one observes "performances" but one cannot safely infer what thought processes produced those performances, and differences in underlying thought processes, perhaps temporarily concealed, are only too likely to irrupt into prominent display at some time in the future.

NOTES

1. Our analysis has consistently made use of these "key sub-goals" -- the idea of joining up "sections" of a proof or of a calculation. In early 1982 we learned of work being done independently in Santa Cruz, California, by Kristina Hooper and her co-workers, and found that they use essentially this same system of analysis. There two independent efforts provide some confirmation of correctness in a field where neither logic nor statistics can usually be used to evaluate conclusions.

2. At the high school level, algebra is typically very different from geometry. For the sequence

$$\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \ldots$$

the nth term can be written as

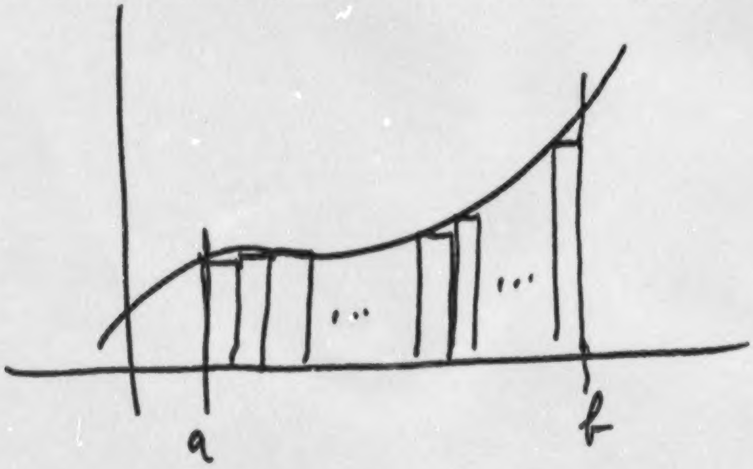
$$\frac{1}{2^n} ,$$

which is a perfectly satisfactory way to write the general term. Similarly,

$$ax^2 + bx + c = 0 \quad a \neq 0$$

is the general quadratic equation.

One cannot do as well with "the general triangle." Whatever triangle one draws has many specific features that are not general at all. This causes difficulty for any student who has not learned the mathematician's common convention of omitting specifics from certain diagrams, as in



where drawing every rectangle would be misleading, since it would seem to show the correct number of rectangles.

3. Clearly, this composite "protocol" is presented here in a considerably simplified form.

4. In fact, two identical numbers would almost certainly be recognized in the first step of the comparison. We have not bothered to indicate this test for identity, which probably occurs at the start of the procedure. The fact that humans probably *begin* with checks for identity, whereas typical computer programs perhaps do not, is food for thought!

5. It is important to note that *function* is a special term in mathematics, and is a noun, not a verb. Within mathematics, there are disagreements in the best way to define *function*, and the historical evolution of the idea may or may not be well represented in certain current definitions. But in any case, "function" is a special term, as in "y is a function of x", or $y = f(x)$.

6. Insertions in square brackets have been added.

7. The importance of "welding together" separate partial representations appears also in recent unpublished work by Kristina Hooper.

8. This lesson is part of the "Madison Project" curriculum, and has consequently been taught to many thousands of students at the 4th, 5th, and 6th grade levels. This precludes precise counting of students -- which would in any case NOT be very revealing, since the proportions of different responses can be altered by minor changes in presentation -- but there is a kind of massive stability to the data, nonetheless.

BIBLIOGRAPHY

ALDERMAN, DONALD L., Spencer S. Swinton and James S. Braswell. "Assessing Basic Arithmetic Skills and Understanding Across Curricula: Computer-Assisted Instruction and Compensatory Education." The Journal of Children's Mathematical Behavior, vol. 2, no. 2 (Spring 1979), pp. 3-28.

ANTON, Howard. Calculus with Analytic Geometry. New York: John Wiley and Sons. 1980.

BLITZ, LEO. "Giant Molecular-Cloud Complexes in the Galaxy." Scientific American, vol. 246, no. 4 (April, 1982), pp. 84-94.

BRANSFORD, John D. and Marcia K. Johnson. "Contextual Prerequisities for Understanding: Some Investigations of Comprehension and Recall." Journal of Verbal Learning and Verbal Behavior, vol. 11 (1972) pp. 717-726.

BROWNELL, William A. "When is Arithmetic Meaningful?" Journal of Educational Research, vol. 38, no. 7 (March, 1945).

BROWNELL, William A, and Verner Sims. "The Nature of Understanding." 45th Yearbook of the National Society for the Study of Education. Chicago: University of Chicago Press, 1946.

BYERS, V. AND N. Herscovics. "Understanding School Mathematics." Mathematics Teaching, vol. 81 (December, 1977), pp. 24-27.

CLEMENT, John and James J. Kaput. "Letter to the Editor." The Journal of Children's Mathematical Behavior, vol. 2, no. 2 (Spring, 1979), p. 208.

CLEMENT, John and Peter Rosnick. "Learning without Understanding: The Effect of Tutoring Strategies on Algebra Misconceptions." The Journal of Mathematical Behavior, vol. 3, no. 1 (Autumn,1980)pp. 3-27.

DAVIS, Robert B. Explorations in Mathematics. A Text for Teachers. Palo Alto, CA: Addison-Wesley, 1967.

DAVIS, Robert B. "Representing Knowledge about Mathematics for Computer-Aided Teaching, Part I. -- Educational Applications of Conceptualizations from Artificial Intelligence." In: E. W. Elcock and Donald Michie (Eds.) Machine Intelligence 8. Machine Representations of Knowledge. Ellis Horwood Ltd. (Distributed by Halstead Press, John Wiley and Sons, Inc., New York) 1977, pp. 363-386.

DAVIS, Robert B. Discovery in Mathematics: A Text for Teachers. New Rochelle, NY: Cuisenaire Company of America, 1980-A.

DAVIS, Robert B. "The Postulation of Certain Specific, Explicit, Commonly-Shared Frames." The Journal of Mathematical Behavior, vol. 3, no. 1 (Autumn 1980 -B), pp. 167-201.

DAVIS, Robert B. "Complex Mathematical Cognition." In: Herbert Ginsburg (Ed.) The Development of Mathematical Thinking. Academic Press 1982-A (in press).

DAVIS, Robert B. Learning Mathematics: The Cognitive Science Approach to Mathematics Education. London: Croom-Helm, 1982-B (in press).

DAVIS, Robert B., Elizabeth Jockusch, and Curtis McKnight. "Cognitive Processes in Learning Algebra." The Journal of Children's Mathematical Behavior, vol. 2, no. 1 (Spring, 1978), pp. 10-320.

DAVIS, Robert B. and Curtis McKnight. "Modeling the Processes of Mathematical Thinking." The Journal of Children's Mathematical Behavior, vol. 2, no. 2 (Spring, 1979), pp. 91-113.

DAVIS, Robert B. and Curtis McKnight. "The Influence of Semantic Content on Algorithmic Behavior. The Journal of Mathematical Behavior, vol. 3, no. 1 (Autumn, 1980), pp. 39-79.

DAVIS, R. and J. King. "An Overview of Production Systems." In: E. W. Elcock and Donald Michie (Eds.) Machine Intelligence 8. Machine Representation of Knowledge. Ellis Horwood Ltd. (Dist. by Halstead Press, John Wiley and Songs, Inc., New York) 1977.

ERLWANGER, Stanley H. "Benny's Conception of Rules and Answers in IPI Mathematics." The Journal of Children's Mathematical Behavior, vol. 1, no. 2 (Autumn, 1973), pp. 7-26.

ERLWANGER, Stanley H. Case Studies of Children's Conceptions of Mathematics. Doctoral Dissertation, University of Illinois, Urbana, Illinois, 1974.

FINIZIO, N. AND G. Ladas. Ordinary Differential Equations with Modern Applications. Belmont, CA: Wadsworth Publishing Company,, 1982.

FRIEND, Jamesine. "Column Addition Skills." The Journal of Children's Mathematical Behavior, vol. 2, no. 2 (Spring, 1979), pp. 29-54.

GINSBURG, Herbert. "The Psychology of Arithmetic Thinking." The Journal of Children's Mathematical Behavior, vol. 1, no. 4 (Spring, 1977), pp. 1-89.

GINSBURG, Herbert. "The Addition Methods of First and Second-Grade Children. Journal for Research in Mathematics Education, vol. 12 (1981),pp.95-106.

HILDEBRAND, F. B. Advanced Calculus for Engineers. New York: Prentice-Hall, Inc., 1949.

HINESLEY, Dan, John Hayes and Herbert Simon. "From Words to Equations: Meaning and Representation in Algebra Word Problems." In: Marcel Just and Patricia Carpenter (Eds.) *Cognitive Processes in Comprehension*. Hillsdale, NJ: Erlbaum, 1977.

HOFSTADTER, Douglas R. *Gödel, Escher and Bach: An Eternal Golden Braid*. New York, NY: Vintage Books, 1980. (Originally published by Basic Books in 1979).

KUMAR, Derek. "Sample Solutions." *The Journal of Mathematical Behavior*, vol. 3, no. 1 (Autumn 1980), pp. 183-199.

LOCHHEAD, Jack. "Faculty Interpretations of Simple Algebraic Statements: The Professor's Side of the Equation." *The Journal of Mathematical Behavior*, vol. 3, no. 1 (Autumn,1980), pp. 29-37.

MATZ, Marilyn. "Towards a Computational Model of Algebraic Competence." *The Journal of Mathematical Behavior*, vol. 3, no. 1 (Autumn, 1980), pp. 93-166.

MINSKY, Marvin. "A Framework for Representing Knowledge." In: P. Winston (Ed.) *The Psychology of Computer Vision*. New York: McGraw-Hill, 1975.

MINSKY, Marvin. "K-Lines: A Theory of Memory." *Cognitive Science*, vol. 4, no. 2 (April-June, 1980), pp. 117-133.

MINSKY, Marvin and Seymour Papert. *Artificial Intelligence Memo No. 252*. Cambridge, MA: Massachusetts Institute of Technology Artificial Intelligence Laboratory, January 1, 1972.

NEWELL, Allen and Herbert A. Simon. *Human Problem Solving*. Englewood Cliffs, NJ: Prentice-Hall, Inc., 1972.

OSGOOD, C. E. *Method and Theory in Experimental Psychology*. Oxford University Press, 1953.

PARKER, Philip. "A General Method for Finding Tangents." *The Journal of Mathematical Behavior*, vol. 3, no. 1 (Autumn, 1980) pp. 208-209.

RISSLAND, Edwina. *The Structure of Mathematical Knowledge*. Cambridge, MA: M.I.T., Artificial Intelligence Laboratory, Technical Report No. 472 August, 1978-A.

RISSLAND, Edwina. "Understanding Understanding Mathematics." *Cognitive Science*, vol. 2 (1978-B), pp. 361-383.

ROSNICK, Peter and John Clement. "Learning Without Understanding: The Effect of Tutoring Strategies on Algebra Misconceptions." *The Journal of Mathematical Behavior*, vol. 3, no. 1 (Autumn, 1980),pp. 3-27.

SHULMAN, Lee S. and Evan R. Keisler. (Eds.) *Learning by Discovery*. Rand McNally, 1966.

SKEMP, Richard R. Intelligence, Learning, and Action. New York, NY: John Wiley and Sons

STEIN, Sherman K. and Calvin D. Crabill. Elementary Algebra: A Guided Inquiry. Boston: Houghton-Mifflin, 1972.

SUZUKI, Kazuko. "Solutions to Problems." The Journal of Children's Mathematical Behavior, vol. 2, no. 2 (Spring 1979), pp. 159-163.

SWINTON, Spencer S., Marianne Amarel, and Judith A. Morgan. The PLATO Elementary Demonstration -- Educational Outcome Evaluation. Final Report: Summary and Conclusions. Princeton, NJ: Educational Testing Service, 1978..

WINOGRAD, Terry. Procedures as a Representation for Data in a Computer Program for Understanding Natural Languages. Ph.D. Thesis. Cambridge, MA; Massachusetts Institute of Technology, Artificial Intelligence Laboratory, 1971.

YOUNG, Stephen C. "The Mental Representation of Geometrical Knowledge." The Journal of Mathematical Behavior, vol. 3, no. 2 (in press).